

Exhibit A  
Application No. 08/587,895

08/587895

METHODS AND COMPOSITIONS FOR IDENTIFYING

RECEPTOR EFFECTORS



### Background Of The Invention

A common technique for cloning receptors is to use nucleic acid hybridization technology to identify receptors which are homologous to other, known receptors. For instance, originally the cloning of seven transmembrane domain G protein-coupled receptors (GCR) depended on the isolation and sequencing of the corresponding protein or the use of expression cloning techniques. However, when sequences for these receptors became available, it was apparent that there were significant sequence homologies between these receptors. This technology, since it does not require that the ligand of the receptor have been identified, has resulted in the cloning of a large number of "orphan receptors", which have no known ligand and often whose biological function is obscure. Receptors of all types comprise this large family. Known orphan receptors include the nuclear receptors COUP-TF1/EAR3, COUP-TF2/ARF1, EAR-1, EAR-2, TR-2, PPAR1, HNF-4, ERF-1, ERF-2, NGFIB/Nur77, ELP/SF-1 and MPL (Parker et al, supra, and Power et al. (1992) TIBS 13:318-323). A large number of orphan receptors have been identified in the EPH family (Hirai et al (1987) *Science* 238:1717-1720). HER3 and HER4 are orphan receptors in the epidermal growth factor receptor family (Flowman et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:1746-1750). ILA is a newly identified member of the human nerve growth factor/tumor necrosis factor receptor family (Schwarz et al. (1993) *Gene* 134:295-298). IRRR is an orphan insulin receptor-related receptor which is a transmembrane tyrosine kinase (Shier et al. (1989) *J. Biol. Chem* 264:14606-14608). Several orphan tyrosine kinase receptors have been found in *Drosophila* (Perrimon (1994) *Curr. Opin. Cell Biol.* 6:260-266). The importance of identifying ligands for orphan receptors is clear; it opens up a wide area for research in the area of drug discovery.

One large subgroup of orphan receptors, as alluded to above, are found in the G protein coupled receptor family. Approximately 100 such receptors have been identified by function and these mediate transmembrane signaling from external stimuli (vision, taste and smell), endocrine function (pituitary and adrenal), exocrine function (pancreas), heart rate,

lipolysis, and carbohydrate metabolism. Structural and genetic similarities suggest that G protein-coupled receptor superfamily can be subclassified into five distinct groups: (i) amine receptors (serotonin, adrenergic, etc.); (ii) small peptide hormone (somatostatin, TRH, etc.); (iii) large peptide hormone (LH-CG, FSH, etc.); (iv) secretin family; and (v) odorant receptors (Buck L. and Axel, R. (1991) *Cell* 65:175-187), with orphan receptors apparently occurring in each of the sub-families.

Previous work describes the expression of recombinant mammalian G protein-coupled receptors as a means of studying receptor function as a means of identifying agonists and antagonists of those receptors. For example, the human muscarinic receptor (HMI) has been functionally expressed in mouse cells (Harbold et al. US Pat. 5,401,629). The rat V1b vasopressin receptor has been found to stimulate phospholipid hydrolysis and intracellular  $Ca^{2+}$  mobilization in Chinese hamster ovary cells upon agonist stimulation (Lolait et al. (1995) *Proc Natl. Acad Sci USA* 92:6783-6787). Likewise, the C5a receptor (to be completed) These types of ectopic expression studies have enabled researchers to study receptor signalling mechanisms and to perform mutagenesis studies which have been useful in identifying portions of receptors that are critical for ligand binding or signal transduction.

Experiments have also been undertaken to express functional G protein coupled receptors in yeast cells. For example, U.S. Patent 5,482,835 to King et al. describes a transformed yeast cell which is incapable of producing a yeast G protein  $\alpha$  subunit, but which has been engineered to produce both a mammalian G protein  $\alpha$ -subunit and a mammalian receptor which is "coupled to" (i.e., interacts with) the aforementioned mammalian G protein  $\alpha$ -subunit. Specifically, U.S. Patent 5,482,835 reports expression of the human beta-2 adrenergic receptor ( $\beta 2AR$ ), a seven transmembrane receptor (STR), in yeast, under control of the GAL1 promoter, with the  $\beta 2AR$  gene modified by replacing the first 63 base pairs of coding sequence with 11 base pairs of noncoding and 42 base pairs of coding sequence from the STE2 gene. (STE2 encodes the yeast  $\alpha$ -factor receptor). The Duke researchers found that the modified  $\beta 2AR$  was functionally integrated into the membrane, as shown by studies of the ability of isolated membranes to interact properly with various known agonists and antagonists of  $\beta 2AR$ . The ligand binding affinity for yeast-expressed  $\beta 2AR$  was said to be nearly identical to that observed for naturally produced  $\beta 2AR$ .

U.S. Patent 5,482,835 describes co-expression of a rat G protein  $\alpha$ -subunit in the same cells, yeast strain 8C, which lacks the cognate yeast protein. Ligand binding resulted in G protein-mediated signal transduction. U.S. Patent 5,482,835 teaches that these cells may be used in screening compounds for the ability to affect the rate of dissociation of G $\alpha$  from G $\beta\gamma$  in a cell. For this purpose, the cell further contains a pheromone-responsive promoter (e.g. BARI or FUS1), linked to an indicator gene (e.g. HIS3 or LacZ). The cells are placed in multi-titer plates, and different compounds are placed in each well. The colonies are then

scored for expression of the indicator gene.

### *Summary of the Invention*

The present invention relates to a rapid, reliable and effective assay for screening and identifying pharmaceutically effective compounds that specifically interact with and modulate the activity of a cellular receptor or ion channel. The subject assay enables rapid screening of large numbers of polypeptides in a library to identify those polypeptides which agonize or antagonize receptor bioactivity. In general, the assay is characterized by the use of a library of recombinant cells, each cell of which include (i) a target receptor protein whose signal transduction activity can be modulated by interaction with an extracellular signal, the transduction activity being able to generate a detectable signal, and (ii) an expressible recombinant gene encoding an exogenous test polypeptide from a polypeptide library. By the use of a variegated gene library, the mixture of cells collectively express a variegated population of test polypeptides. In preferred embodiments, the polypeptide library includes at least  $10^3$  different polypeptides, though more preferably at least  $10^5$ ,  $10^6$ , or  $10^7$  different (variegated) polypeptides. The polypeptide library can be generated as a random peptide library, as a semi-random peptide library (e.g., based on combinatorial mutagenesis of a known ligand), or as a cDNA library.

The ability of particular constituents of the peptide library to modulate the signal transduction activity of the target receptor can be scored for by detecting up or down-regulation of the detection signal. For example, second messenger generation via the receptor can be measured directly. Alternatively, the use of a reporter gene can provide a convenient readout. In any event, a statistically significant change in the detection signal can be used to facilitate isolation of those cells from the mixture which contain a nucleic acid encoding a test polypeptide which is an effector of the target receptor.

By this method, test polypeptides which induce receptor signaling can be identified. If the test polypeptide does not appear to directly induce the activity of the receptor protein, the assay may be repeated and modified by the introduction of a step in which the recombinant cell is first contacted with a known activator of the target receptor to induce the signal transduction pathways from the receptor. In one embodiment, the test polypeptide is assayed for its ability to antagonize, e.g., inhibit or block the activity of the activator. Alternatively, the assay can score for peptides from the peptide library which potentiate the induction response generated by treatment of the cell with a known activator. As used herein, an "agonist" refers to agents which either induce activation of receptor signaling pathways, e.g., such as by mimicking a ligand for the receptor, as well as agents which potentiate the sensitivity of the receptor to a ligand, e.g., lower the concentrations of ligand required to induce a particular level of receptor-dependent signaling.



The receptor protein can be any receptor which interacts with an extracellular molecule (i.e. hormone, growth factor, peptide) to modulate a signal in the cell. To illustrate the receptor can be a cell surface receptor, or in other embodiments can be an intracellular receptor. In preferred embodiments, the receptor is a cell surface receptor, such as: a receptor tyrosine kinase, e.g., an EPH receptor; an ion channel; a cytokine receptor; an multisubunit immune recognition receptor; a chemokine receptor; a growth factor receptor, or a G-protein coupled receptor, such as a chemoattractant peptide receptor, a neuropeptide receptor, a light receptor, a neurotransmitter receptor, or a polypeptide hormone receptor.

- 4 -

receptor, antidiuretic hormone receptor, glucagon receptor, and adrenocorticotrophic hormone receptor.

Preferred EPH receptors include *eph, elk, eck, sek, mek4, hek, hek2, eek, erk, tyro1, tyro4, tyro5, tyro6, tyro11, cek4, cek5, cek6, cek7, cek8, cek9, cek10, bsk, rik1, rik2, rik3, myk1, myk2, ehk1, ehk2, pagliaccio, hik, erk* and *nuk* receptors.

As set forth below, no matter which structural/function class to which the target receptor may belong, the subject assay is amenable to identifying ligands for an otherwise orphan receptor.

In those embodiments wherein the target receptor is a cell surface receptor, it will be desirable for the peptides in the library to express a signal sequence to ensure that they are processed in the appropriate secretory pathway and thus are available to interact with receptors on the cell surface.

With respect to a detection signal generated by signal transduction, certain of the preferred embodiments measure the production of second messengers to determine changes in ligand engagement by the receptor. In preferred embodiments, changes in GTP hydrolysis, calcium mobilization, or phospholipid hydrolysis can be measured.

In other preferred embodiment, the host cells harbors a reporter construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the receptor protein. Exemplary reporter genes include enzymes, such as luciferase, phosphatase, or  $\beta$ -galactosidase which can produce a spectrometrically active label, e.g., changes in color, fluorescence or luminescence, or a gene product which alters a cellular phenotype, e.g., cell growth, drug resistance or auxotrophy. In preferred embodiments: the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase,  $\beta$ -galactosidase and secreted alkaline phosphatase; the reporter gene encodes a gene product which confers a growth signal; the reporter gene encodes a gene product for growth in media containing aminotriazole or canavanine.

The reagent cells of the present invention can be derived from any eukaryotic organism. In preferred embodiments the cells are mammalian cells. In more preferred embodiments the cells are yeast cells, with cells from the genera *Saccharomyces* or *Schizosaccharomyces* being more preferred. However, cells from amphibia (such as xenopus), avian or insect sources are also contemplated. The host cells can derived from primary cells, or transformed and/or immortalized cell lines.

### *Brief Description of the Drawings*

Figure 1. Structures of PAAH5 and PRS-ADC.

Figure 2. Schematic diagram of the structure of the plasmid resulting from insertion of random oligonucleotides into PADC-MF alpha. This plasmid expresses random peptides in the context of the MF alpha 1 signal and leader peptide.

Figure 3. Schematic diagram of the structure of the plasmid resulting from insertion of random oligonucleotides into PADC-MF alpha. This plasmid expresses random peptides in the context of the MF alpha 1 leader and C-terminal CVIA tetrapeptide.

Figure 4. Activity of a fus1 promoter in response to signaling by human C5a expressed in autocrine strains of yeast.

#### *Detailed Description of the Invention*

Proliferation, differentiation and death of eukaryotic cells are controlled by hormones, neurotransmitters, and polypeptide factors. These diffusible ligands allow cells to influence and be influenced by environmental cues. The study of receptor-ligand interaction has revealed a great deal of information about how cells respond to external stimuli, and this knowledge has led to the development of therapeutically important compounds. However, the rate at which receptors have been cloned has recently increased dramatically -- existing families have been extended and new families recognized. In particular, the application of advanced cloning approaches has allowed the isolation of many receptors for which ligands are initially unknown. These are commonly referred to in the art as "orphan" receptors, and several have subsequently proved to be important pharmacological targets.

The present invention makes available a rapid, effective assay for screening and identifying pharmacologically effective compounds that specifically interact with and modulate the activity of a cellular receptor or ion channel. The subject assay enables rapid screening of large numbers of polypeptides in a library to identifying those polypeptides which induce or antagonize receptor bioactivity.

In general, the assay is characterized by the use of a mixture of recombinant cells to sample a variegated polypeptide library for receptor agonists or antagonists. As described with greater detail below, the reagent cells express both a target receptor protein capable of transducing a detectable signal in the reagent cell, and a test polypeptide for which interaction with the receptor is to be ascertained. Collectively, a culture of such reagent cells will provide a variegated library of potential receptor effectors and those members of the library which either agonize or antagonize the receptor function can be selected and identified by sequence.

One salient feature of the subject assay is the enhanced sensitivity resulting from

expression of the test polypeptide in a cell which also serves as a reporter for the desired receptor-ligand interaction. To illustrate, where the detectable signal resulting from receptor engagement by an agonist provides a growth signal or drug resistance, individual cells expressing polypeptides which agonize receptor function can be amplified and isolated from a library culture.

Accordingly, the present invention provides a convenient format for discovering drugs which can be useful to modulate cellular function, as well as to understand the pharmacology of compounds that specifically interact with cellular receptors or ion channels. Moreover, the subject assay is particularly amenable to identifying ligands, natural or artificial, for orphan receptors.

Before further description of the invention, certain terms employed in the specification, examples and appended claims are, for convenience, collected here.

As used herein, "recombinant cells" include any cells that have been modified by the introduction of heterologous DNA. Control cells include cells that are substantially identical to the recombinant cells, but do not express one or more of the proteins encoded by the heterologous DNA, e.g., do not include or express the reporter gene construct, receptor or test polypeptide.

The terms "recombinant protein", "heterologous protein" and "exogenous protein" are used interchangeably throughout the specification and refer to a polypeptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous protein. That is, the polypeptide is expressed from a heterologous nucleic acid.

As used herein, "heterologous DNA" or "heterologous nucleic acid" include DNA that does not occur naturally as part of the genome in which it is present or which is found in a location or locations in the genome that differs from that in which it occurs in nature. Heterologous DNA is not endogenous to the cell into which it is introduced, but has been obtained from another cell. Generally, although not necessarily, such DNA encodes RNA and proteins that are not normally produced by the cell in which it is expressed. Heterologous DNA may also be referred to as foreign DNA. Any DNA that one of skill in the art would recognize or consider as heterologous or foreign to the cell in which is expressed is herein encompassed by heterologous DNA. Examples of heterologous DNA include, but are not limited to, DNA that encodes test polypeptides, receptors, reporter genes, transcriptional and translational regulatory sequences, selectable or traceable marker proteins, such as a protein that confers drug resistance.

As used herein, "cell surface receptor" refers to molecules that occur on the surface of cells, interact with the extracellular environment, and transmit or transduce the information

regarding the environment intracellularly in a manner that ultimately modulates transcription of specific promoters, resulting in transcription of specific genes.

As used herein, "extracellular signals" include a molecule or a change in the environment that is transduced intracellularly via cell surface proteins that interact, directly or indirectly, with the signal. An extracellular signal or effector molecule includes any compound or substance that in some manner specifically alters the activity of a cell surface protein. Examples of such signals include, but are not limited to, molecules such as acetylcholine, growth factors and hormones, that bind to cell surface and/or intracellular receptors and ion channels and modulate the activity of such receptors and channels.

As used herein, "extracellular signals" also include as yet unidentified substances that modulate the activity of a cellular receptor, and thereby influence intracellular functions. Such extracellular signals are potential pharmacological agents that may be used to treat specific diseases by modulating the activity of specific cell surface receptors.

"Orphan receptors" is a designation given to a receptors for which no specific natural ligand has been described.

As used herein, a "reporter gene construct" is a nucleic acid that includes a "reporter gene" operatively linked to a transcriptional regulatory sequences. Transcription of the reporter gene is controlled by these sequences. The activity of at least one or more of these control sequences is directly or indirectly regulated by the target receptor protein. The transcriptional regulatory sequences include the promoter and other regulatory regions, such as enhancer sequences, that modulate the activity of the promoter, or regulatory sequences that modulate the activity or efficiency of the RNA polymerase that recognizes the promoter, or regulatory sequences are recognized by effector molecules, including those that are specifically induced by interaction of an extracellular signal with the target receptor. For example, modulation of the activity of the promoter may be effected by altering the RNA polymerase binding to the promoter region, or, alternatively, by interfering with initiation of transcription or elongation of the mRNA. Such sequences are herein collectively referred to as transcriptional regulatory elements or sequences. In addition, the construct may include sequences of nucleotides that alter translation of the resulting mRNA, thereby altering the amount of reporter gene product.

"Signal transduction" is the processing of chemical signals from the cellular environment through the cell membrane, and may occur through one or more of several mechanisms, such as phosphorylation, activation of ion channels, effector enzyme activation via guanine nucleotide binding protein intermediates, formation of inositol phosphate, activation of adenylyl cyclase, and/or direct activation (or inhibition) of a transcriptional factor.

The term "modulation of a signal transduction activity of a receptor protein" in its

various grammatical forms, as used herein, designates induction and/or potentiation, as well as inhibition of one or more signal transduction pathways downstream of a receptor.

Agonists and antagonists are "receptor effector" molecules that modulate signal transduction via a receptor. Receptor effector molecules are capable of binding to the receptor, though not necessarily at the binding site of the natural ligand. Receptor effectors can modulate signal transduction when used alone, i.e. can be surrogate ligands, or can alter signal transduction in the presence of the natural ligand, either to enhance or inhibit signaling by the natural ligand. For example, "antagonists" are molecules that block or decrease the signal transduction activity of receptor, e.g., they can competitively, noncompetitively, and/or allosterically inhibit signal transduction from the receptor, whereas "agonists" potentiate, induce or otherwise enhance the signal transduction activity of a receptor. The terms "receptor activator" and "surrogate ligand" refer to an agonist which induces signal transduction from a receptor.

The term "substantially homologous", when used in connection with amino acid sequences, refers to sequences which are substantially identical to or similar in sequence, giving rise to a homology in conformation and thus to similar biological activity. The term is not intended to imply a common evolution of the sequences.

Typically, "substantially homologous" sequences are at least 50%, more preferably at least 80%, identical in sequence, at least over any regions known to be involved in the desired activity. Most preferably, no more than five residues, other than at the termini, are different. Preferably, the divergence in sequence, at least in the aforementioned regions, is in the form of "conservative modifications".

The term "autocrine cell", as used herein, refers to a cell which produces a substance which can stimulate a receptor located on or within the same cell as produces the substance. For example, wild-type yeast  $\alpha$  and  $\alpha$  cells are not autocrine. However, a yeast cell which produces both  $\alpha$ -factor and  $\alpha$ -factor receptor, or both  $\alpha$ -factor and  $\alpha$ -factor receptor, in functional form, is autocrine. By extension, cells which produce a peptide which is being screened for the ability to activate a receptor (e.g., by activating a G protein-coupled receptor) express the receptor are called "autocrine cells", though it might be more precise to call them "putative autocrine cells". Of course, in a library of such cells, in which a multitude of different peptides are produced, it is likely that one or more of the cells will be "autocrine" in the stricter sense of the term.

The terms "protein", "polypeptide" and "peptide" are used interchangeably herein.

## *1. Overview of Assay*

As set out above, the present invention relates to methods for identifying effectors of

a receptor protein or complex thereof. In general, the assay is characterized by the use of a library of recombinant cells, each cell of which include (i) a target receptor protein whose signal transduction activity can be modulated by interaction with an extracellular signal, the transduction activity being able to generate a detectable signal, and (ii) an expressible recombinant gene encoding an exogenous test polypeptide from a polypeptide library. By the use of a variegated gene library, the mixture of cells collectively express a variegated population of test polypeptides.

The ability of particular constituents of the peptide library to modulate the signal transduction activity of the target receptor can be scored for by detecting up or down-regulation of the detection signal. For example, second messenger generation (e.g. GTPase activity, phospholipid hydrolysis, or protein phosphorylation) via the receptor can be measured directly. Alternatively, the use of a reporter gene can provide a convenient readout. In any event, a statistically significant change in the detection signal can be used to facilitate isolation of those cells from the mixture which contain a nucleic acid encoding a test polypeptide which is an effector of the target receptor.

By this method, test polypeptides which induce the receptor's signaling can be screened. If the test polypeptide does not appear to induce the activity of the receptor protein, the assay may be repeated and modified by the introduction of a step in which the recombinant cell is first contacted with a known activator of the target receptor to induce signal transduction from the receptor, and the test polypeptide is assayed for its ability to inhibit the activity of the receptor, e.g., to identify receptor antagonists. In yet other embodiments, the peptide library can be screened for members which potentiate the response to a known activator of the receptor. In this respect, surrogate ligands identified by the present assay for orphan receptors can be used as the exogenous activator, and further peptide libraries screened for members which potentiate or inhibit the activating peptide. Alternatively, the surrogate ligand can be used to screen exogenous compound libraries (peptide and non-peptide) which, by modulating the activity of the identified surrogate, will presumably also similarly effect the native ligand's effect on the target receptor. In such embodiments, the surrogate ligand can be applied to the cells, though is preferably produced by the reagent cell, thereby providing an autocrine cell.

In developing the recombinant cells assays, it was recognized that a frequent result of receptor-mediated responses to extracellular signals was the transcriptional activation or inactivation of specific genes after exposure of the cognate receptor to an extracellular signal that induces such activity. Thus, transcription of genes controlled by receptor-responsive transcriptional elements often reflects the activity of the surface protein by virtue of transduction of an intracellular signal.

To illustrate, the intracellular signal that is transduced can be initiated by the specific

interaction of an extracellular signal, particularly a ligand, with a cell surface receptor on the cell. This interaction sets in motion a cascade of intracellular events, the ultimate consequence of which is a rapid and detectable change in the transcription or translation of a gene. By selecting transcriptional regulatory sequences that are responsive to the transduced intracellular signals and operatively linking the selected promoters to reporter genes, whose transcription, translation or ultimate activity is readily detectable and measurable, the transcription based assay provides a rapid indication of whether a specific receptor or ion channel interacts with a test peptide in any way that influences intracellular transduction. Expression of the reporter gene, thus, provides a valuable screening tool for the development of compounds that act as agonists or antagonists of a cell receptor or ion channel.

Reporter gene based assays of this invention measure the end stage of the above described cascade of events, e.g., transcriptional modulation. Accordingly, in practicing one embodiment of the assay, a reporter gene construct is inserted into the reagent cell in order to generate a detection signal dependent on receptor signaling. Typically, the reporter gene construct will include a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the target receptor, with the level of expression of the reporter gene providing the receptor-dependent detection signal. The amount of transcription from the reporter gene may be measured using any method known to those of skill in the art to be suitable. For example, specific mRNA expression may be detected using Northern blots or specific protein product may be identified by a characteristic stain or an intrinsic activity.

In preferred embodiments, the gene product of the reporter is detected by an intrinsic activity associated with that product. For instance, the reporter gene may encode a gene product that, by enzymatic activity, gives rise to a detection signal based on color, fluorescence, or luminescence.

The amount of expression from the reporter gene is then compared to the amount of expression in either the same cell in the absence of the test compound or it may be compared with the amount of transcription in a substantially identical cell that lacks the specific receptors. A control cell may be derived from the same cells from which the recombinant cell was prepared but which had not been modified by introduction of heterologous DNA, e.g., the encoding the test polypeptide. Alternatively, it may be a cell in which the specific receptors are removed. Any statistically or otherwise significant difference in the amount of transcription indicates that the test polypeptide has in some manner altered the activity of the specific receptor.

In other preferred embodiments, the reporter or marker gene provides a selection method such that cells in which the peptide is a ligand for the receptor have a growth advantage. For example the reporter could enhance cell viability, relieve a cell nutritional



requirement, and/or provide resistance to a drug.

With respect to the target receptor, it may be endogenously expressed by the host cell, or it may be expressed from a heterologous gene that has been introduced into the cell. Methods for introducing heterologous DNA into eukaryotic cells are of course well known in the art and any such method may be used. In addition, DNA encoding various receptor proteins is known to those of skill in the art or it may be cloned by any method known to those of skill in the art. In certain embodiments, such as when an exogenous receptor is expressed, it may be desirable to inactivate, such as by deletion, a homologous receptor present in the cell.

The subject assay is useful for identifying polypeptides that interact with any receptor protein whose activity ultimately induces a signal transduction cascade in the host cell which can be exploited to produce a detectable signal. In particular, the assays can be used to test functional ligand-receptor or ligand-ion channel interactions for cell surface-localized receptors and channels, and also for cytoplasmic and nuclear receptors. As described in more detail below, the subject assay can be used to identify effectors of, for example, G protein-coupled receptors, receptor tyrosine kinases, cytokine receptors, and ion channels, as well as steroid hormone receptors. In preferred embodiments the method described herein is used for identifying ligands for "orphan receptors" for which no ligand is known.

In embodiments in which cell surface receptors are the assay targets, it will be desirable for each of the peptides of the peptide library to include a signal sequence for secretion, e.g., which will ensure appropriate transport of the peptide to the endoplasmic reticulum, the golgi, and ultimately to the cell surface so that it is able to interact with cell surface receptors. In the case of yeast cells, the signal sequence will transport peptides to the periplasmic space.

Any transfectable cell that can express the desired cell surface protein in a manner such that the protein functions to intracellularly transduce an extracellular signal may be used. The cells may be selected such that they endogenously express the target receptor protein or may be genetically engineered to do so.

The preparation of cells which express the orphan FRL1 receptor, a peptide library, and a reporter gene expression construct, are described. These cells have been used to identify a novel ligand for this receptor. The cells for the identification of receptor ligands and in drug screening assays to discover agents capable of modulating receptor activity.

Any cell surface protein that is known to those of skill in the art or that may be identified by those of skill in the art may be used in the assay. The cell surface protein may be endogenously expressed on the selected cell or it may be expressed from cloned DNA.

## II. Host Cells

Suitable host cells for generating the subject assay include prokaryotes, yeast, or higher eukaryotic cells, especially mammalian cells. Prokaryotes include gram negative or gram positive organisms. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651) (Gluzman (1981) Cell 23:175) CV-1 cells (ATCC CCL 70), L cells, C127, 3T3, Chinese hamster ovary (CHO), HeLa and BHK cell lines.

If yeast cells are used, the yeast may be of any species which are cultivable and in which an exogenous receptor can be made to engage the appropriate signal transduction machinery of the host cell. Suitable species include *Kluyverei lactis*, *Schizosaccharomyces pombe*, and *Ustilago maydis*; *Saccharomyces cerevisiae* is preferred. Other yeast which can be used in practicing the present invention are *Neurospora crassa*, *Aspergillus niger*, *Aspergillus nidulans*, *Pichia pastoris*, *Candida tropicalis*, and *Hansenula polymorpha*. The term "yeast", as used herein, includes not only yeast in a strictly taxonomic sense, i.e., unicellular organisms, but also yeast-like multicellular fungi or filamentous fungi.

The choice of appropriate host cell will also be influenced by the choice of detection signal. For instance, reporter constructs, as described below, can provide a selectable or screenable trait upon transcriptional activation (or inactivation) in response to a signal transduction pathway coupled to the target receptor. The reporter gene may be an unmodified gene already in the host cell pathway, such as the genes responsible for growth arrest in yeast. It may be a host cell gene that has been operably linked to a "receptor-responsive" promoter. Alternatively, it may be a heterologous gene that has been so linked. Suitable genes and promoters are discussed below. In other embodiments, second messenger generation can be measured directly in the detection step, such as mobilization of intracellular calcium or phospholipid metabolism are quantitated. Accordingly, it will be understood that to achieve selection or screening, the host cell must have an appropriate phenotype. For example, introducing a pheromone-responsive chimeric HIS3 gene into a yeast that has a wild-type HIS3 gene would frustrate genetic selection. Thus, to achieve nutritional selection, an auxotrophic strain is wanted.

To further illustrate, in a preferred embodiment of the subject assay using a yeast host cell, the yeast cells possess one or more of the following characteristics: (a) the endogenous FUS1 gene has been inactivated; (b) the endogenous SST2 gene, and/or other genes involved in desensitization, has been inactivated; (c) if there is a homologous, endogenous receptor gene it has been inactivated; and (d) if the yeast produces an endogenous ligand to the exogenous receptor, the genes encoding for the ligand have been inactivated.

Other complementations for use in the subject assay can be constructed without any undue experimentation. Indeed, many yeast genetic complementation with mammalian

signal transduction proteins have been described in the art. For example, Mosteller et al. (1994) Mol Cell Biol 14:1104-12 demonstrates that human Ras proteins can complement loss of ras mutations in *S. cerevisiae*. Moreover, Toda et al. (1986) Princess Takamatsu Symp 17: 253-60 have shown that human ras proteins can complement the loss of RAS1 and RAS2 proteins in yeast, and hence are functionally homologous. Both human and yeast RAS proteins can stimulate the magnesium and guanine nucleotide-dependent adenylate cyclase activity present in yeast membranes. Ballester et al. (1989) Cell 59: 681-6 describe a vector to express the mammalian GAP protein in the yeast *S. cerevisiae*. When expressed in yeast, GAP inhibits the function of the human ras protein, and complements the loss of IRA1. IRA1 is a yeast gene that encodes a protein with homology to GAP and acts upstream of RAS. Mammalian GAP can therefore function in yeast and interact with yeast RAS. Wei et al. (1994) Gene 151: 279-84 describes that a human Ras-specific guanine nucleotide-exchange factor, Cdc25GEF, can complement the loss of CDC25 function in *S. cerevisiae*. Martegani et al. (1992) EMBO J 11: 2151-7 describe the cloning by functional complementation of a mouse cDNA encoding a homolog of CDC25, a *Saccharomyces cerevisiae* RAS activator. Vojtek et al. (1993) J Cell Sci 105: 777-85 and Mativ et al. (1992) Mol Cell Biol 12: 5033-40 describe how a mouse CAP protein, e.g., an adenylate cyclase associated protein associated with ras-mediated signal transduction, can complement defects in *S. cerevisiae*. Papasavas et al. (1992) Biochem Biophys Res Commun 184:1378-85 also suggest that inactivated yeast adenylate cyclase can be complemented by a mammalian adenylate cyclase gene. Hughes et al. (1993) Nature 364: 349-52 describe the complementation of byr1 in fission yeast by mammalian MAP kinase (MEK). Parissenti et al. (1993) Mol Cell Endocrinol 98: 9-16 describes the reconstitution of bovine protein kinase C (PKC) in yeast. The Ca(2+)- and phospholipid-dependent Ser/Thr kinase PKC plays important roles in the transduction of cellular signals in mammalian cells. Marcus et al. (1995) PNAS 92: 6180-4 suggests the complementation of shk1 null mutations in *S. pombe* by the either the structurally related *S. cerevisiae* Ste20 or mammalian p65PAK protein kinases.

"Inactivation", with respect to genes of the host cell, means that production of a functional gene product is prevented or inhibited. Inactivation may be achieved by deletion of the gene, mutation of the promoter so that expression does not occur, or mutation of the coding sequence so that the gene product is inactive. Inactivation may be partial or total.

"Complementation", with respect to genes of the host cell, means that at least partial function of inactivated gene of the host cell is supplied by an exogenous nucleic acid. For instance, yeast cells can be "mammalianized", and even "humanized", by complementation of receptor and signal transduction proteins with mammalian homologs. To illustrate, inactivation of a yeast Byr2/Ste11 gene can be complemented by expression of a human MEKK gene.

### III. Expression Systems

Ligating a polynucleotide coding sequence into a gene construct, such as an expression vector, and transforming or transfecting into hosts, either eukaryotic (yeast, avian, insect or mammalian) or prokaryotic (bacterial cells), are standard procedures used in producing other well-known proteins, including sequences encoding exogenous receptor and peptide libraries. Similar procedures, or modifications thereof, can be employed to prepare recombinant reagent cells of the present invention by tissue-culture technology in accord with the subject invention.

In general, it will be desirable that the vector be capable of replication in the host cell. It may be a DNA which is integrated into the host genome, and thereafter is replicated as a part of the chromosomal DNA, or it may be DNA which replicates autonomously, as in the case of a plasmid. In the latter case, the vector will include an origin of replication which is functional in the host. In the case of an integrating vector, the vector may include sequences which facilitate integration, e.g., sequences homologous to host sequences, or encoding integrases.

Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, and mammalian cellular hosts are known in the art, and are described in, for example, Powels et al. (*Cloning Vectors: A Laboratory Manual*, Elsevier, New York, 1985). Mammalian expression vectors may comprise non-transcribed elements such as an origin of replication, a suitable promoter and enhancer linked to the gene to be expressed, and other 5' or 3' flanking nontranscribed sequences, and 5' or 3' untranslated sequences, such as necessary ribosome binding sites, a poly-adenylation site, splice donor and acceptor sites, and transcriptional termination sequences.

The preferred mammalian expression vectors contain both prokaryotic sequences, to facilitate the propagation of the vector in bacteria, and one or more eukaryotic transcription units that are expressed in eukaryotic cells. The pCDNAL/amp, pCDNAL/neo, pRc/CMV, pSV2gpt, pSV2neo, pSV2-dht, pTK2, pRSVneo, pMSG, pSVT7, pko-neo and pHyg derived vectors are examples of mammalian expression vectors suitable for transfection of eukaryotic cells. Some of these vectors are modified with sequences from bacterial plasmids, such as pBR322, to facilitate replication and drug resistance selection in both prokaryotic and eukaryotic cells. Alternatively, derivatives of viruses such as the bovine papillomavirus (BPV-1), or Epstein-Barr virus (PHEBo, pREP-derived and p205) can be used for transient expression of proteins in eukaryotic cells. The various methods employed in the preparation of the plasmids and transformation of host organisms are well known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells, as well as general recombinant procedures, see *Molecular Cloning A Laboratory Manual*, 2nd Ed., ed. by

The transcriptional and translational control sequences in expression vectors to be used in transforming mammalian cells may be provided by viral sources. For example, commonly used promoters and enhancers are derived from Polyoma, Adenovirus 2, Simian Virus 40 (SV40), and human cytomegalovirus. DNA sequences derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites may be used to provide the other genetic elements required for expression of a heterologous DNA sequence. The early and late promoters are particularly useful because both are obtained easily from the virus as a fragment which also contains the SV40 viral origin of replication (Fiers et al.(1978) *Nature* 273:111) Smaller or larger SV40 fragments may also be used, provided the approximately 250 bp sequence extending from the Hind III site toward the Bgl I site located in the viral origin of replication is included. Exemplary vectors can be constructed as disclosed by Okayama and Berg (1983, *Mol. Cell Biol.* 3:280). A useful system for stable high level expression of mammalian receptor cDNAs in C127 murine mammary epithelial cells can be constructed substantially as described by Cosman et al (1986, *Mol. Immunol.* 23:935). Other expression vectors for use in mammalian host cells are derived from retroviruses.

In other embodiments, the use of viral transfection can provide stably integrated copies of the expression construct. In particular, the use of retroviral, adenoviral or adeno-associated viral vectors is contemplated as a means for providing a stably transfected cell line which expresses an exogenous receptor, and/or a polypeptide library.

A number of vectors exist for the expression of recombinant proteins in yeast. For instance, YEP24, YIP5, YEP51, YEP52, pYES2, and YRP17 are cloning and expression vehicles useful in the introduction of genetic constructs into *S. cerevisiae* (see, for example, Broach et al. (1983) in *Experimental Manipulation of Gene Expression*, ed. M. Inouye Academic Press, p. 83, incorporated by reference herein). These vectors can replicate in *E. coli* due to the presence of the pBR322 ori, and in *S. cerevisiae* due to the replication determinant of the yeast 2 micron plasmid. In addition, drug resistance markers such as ampicillin can be used. Moreover, if yeast are used as a host cell, it will be understood that the expression of a gene in a yeast cell requires a promoter which is functional in yeast. Suitable promoters include the promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., *J. Biol. Chem.* 255, 2073 (1980) or other glycolytic enzymes (Hess et al., *J. Adv. Enzyme Reg.* 7, 149 (1968); and Holland et al. *Biochemistry* 17, 4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phospho-fructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phospho-glucose isomerase, and glucokinase. Suitable

vectors and promoters for use in yeast expression are further described in R. Hitzeman et al., EPO Publn. No. 73,657. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytoschrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and the aforementioned metallothionein and glyceraldehyde-3-phosphate dehydrogenase, as well as enzymes responsible for maltose and galactose utilization. Finally, promoters that are active in only one of the two haploid mating types may be appropriate in certain circumstances. Among these haploid-specific promoters, the pheromone promoters Mfal and Mfal are of particular interest.

In some instances, it may be desirable to derive the host cell using insect cells. In such embodiments, recombinant polypeptides can be expressed by the use of a baculovirus expression system. Examples of such baculovirus expression systems include pVL-derived vectors (such as pVL1392, pVL1393 and pVL941), pAcUW-derived vectors (such as pAcUW1), and pBlueBac-derived vectors (such as the  $\beta$ -gal containing pBlueBac III).

Libraries of random peptides or cDNA fragments may be expressed in a multiplicity of ways, including as portions of chimeric proteins. As described below, where secretion of the peptide library is desired, the peptide library can be engineered for secretion or transport to the extracellular space via the yeast pheromone system

In constructing suitable expression plasmids, the termination sequences associated with these genes, or with other genes which are efficiently expressed in yeast, may also be ligated into the expression vector 3' of the heterologous coding sequences to provide polyadenylation and termination of the mRNA.

#### *IV. Periplasmic Secretion*

If yeast cells are used as the host cell it will be noted that the yeast cell is bounded by a lipid bilayer called the plasma membrane. Between this plasma membrane and the cell wall is the periplasmic space. Peptides secreted by yeast cells cross the plasma membrane through a variety of mechanisms and thereby enter the periplasmic space. The secreted peptides are then free to interact with other molecules that are present in the periplasm or displayed on the outer surface of the plasma membrane. The peptides then either undergo re-uptake into the cell, diffuse through the cell wall into the medium, or become degraded within the periplasmic space.

The test polypeptide library may be secreted into the periplasm by any of a number of exemplary mechanisms, depending on the nature of the expression system to which they are linked. In one embodiment, the peptide may be structurally linked to a yeast signal sequence, such as that present in the  $\alpha$ -factor precursor, which directs secretion through the

endoplasmic reticulum and Golgi apparatus. Since this is the same route that the receptor protein follows in its journey to the plasma membrane, opportunity exists in cells expressing both the receptor and the peptide library for a specific peptide to interact with the receptor during transit through the secretory pathway. This has been postulated to occur in mammalian cells exhibiting autocrine activation. Such interaction could yield activation of the response pathway during transit, which would still allow identification of those cells expressing a peptide agonist. For situations in which peptide antagonists to externally applied receptor agonist are sought, this system would still be effective, since both the peptide antagonist and receptor would be delivered to the outside of the cell in concert. Thus, those cells producing an antagonist would be selectable, since the peptide antagonist would be properly and timely situated to prevent the receptor from being stimulated by the externally applied agonist.

An alternative mechanism for delivering peptides to the periplasmic space is to use the ATP-dependent transporters of the STE6/MDR1 class. This transport pathway and the signals that direct a protein or peptide to this pathway are not as well characterized as is the endoplasmic reticulum-based secretory pathway. Nonetheless, these transporters apparently can efficiently export certain peptides directly across the plasma membrane, without the peptides having to transit the ER/Golgi pathway. It is anticipated that at least a subset of peptides can be secreted through this pathway by expressing the library in context of the a-factor prosequence and terminal tetrapeptide. The possible advantage of this system is that the receptor and peptide do not come into contact until both are delivered to the external surface of the cell. Thus, this system strictly mimics the situation of an agonist or antagonist that is normally delivered from outside the cell. Use of either of the described pathways is within the scope of the invention.

The present invention does not require periplasmic secretion, or, if such secretion is provided, any particular secretion signal or transport pathway.

## *V. Cytokine Receptors*

In one embodiment the target receptor is a cytokine receptor. Cytokines are a family of soluble mediators of cell-to-cell communication that includes interleukins, interferons, and colony-stimulating factors. The characteristic features of cytokines lie in their functional redundancy and pleiotropy. Most of the cytokine receptors that constitute distinct superfamilies do not possess intrinsic protein tyrosine kinase domains, yet receptor stimulation usually invokes rapid tyrosine phosphorylation of intracellular proteins, including the receptors themselves. Many members of the cytokine receptor superfamily activate the Jak protein tyrosine kinase family, with resultant phosphorylation of the STAT transcriptional activator factors. IL-2, IL-7, IL-2 and Interferon  $\gamma$  have all been shown to

activate Jak kinases (Frank et al (1995) *Proc Natl Acad Sci USA* 92:7779-7783); Scharfe et al. (1995) *Blood* 86:2077-2085); (Bacon et al. (1995) *Proc Natl Acad Sci USA* 92:7307-7311); and (Sakatsume et al (1995) *J. Biol Chem* 270:17528-17534). Events downstream of Jak phosphorylation have also been elucidated. For example, exposure of T lymphocytes to IL-2 has been shown to lead to the phosphorylation of signal transducers and activators of transcription (STAT) proteins STAT1 $\alpha$ , STAT2 $\beta$ , and STAT3, as well as of two STAT-related proteins, p94 and p95. The STAT proteins were found to translocate to the nucleus and to bind to a specific DNA sequence, thus suggesting a mechanism by which IL-2 may activate specific genes involved in immune cell function (Frank et al. *supra*). Jak3 is associated with the gamma chain of the IL-2, IL-4, and IL-7 cytokine receptors (Fujii et al. (1995) *Proc Natl Acad Sci* 92:5482-5486) and (Musso et al (1995) *J Exp Med*. 181:1425-1431). The Jak kinases have also been shown to be activated by numerous ligands that signal via cytokine receptors such as, growth hormone and erythropoietin and IL-6 (Kishimoto (1994) *Stem cells Suppl* 12:37-44).

Detection signals which may be scored for in the present assay, in addition to direct detection of second messengers, such as by changes in phosphorylation, includes reporter constructs which include transcriptional regulatory elements responsive to the STAT proteins. Described *infra*.

#### VI. Multisubunit Immune Recognition Receptor (MIRR)

In another embodiment the receptor is a multisubunit receptor. Receptors can be comprised of multiple proteins referred to as subunits, one category of which is referred to as a multisubunit receptor is a multisubunit immune recognition receptor (MIRR). MIRRs include receptors having multiple noncovalently associated subunits and are capable of interacting with src-family tyrosine kinases. MIRRs can include, but are not limited to, B cell antigen receptors, T cell antigen receptors, Fc receptors and CD22. One example of an MIRR is an antigen receptor on the surface of a B cell. To further illustrate, the MIRR on the surface of a B cell comprises membrane-bound immunoglobulin (mIg) associated with the subunits Ig- $\alpha$  and Ig- $\beta$  or Ig- $\gamma$ , which forms a complex capable of regulating B cell function when bound by antigen. An antigen receptor can be functionally linked to an amplifier molecule in a manner such that the amplifier molecule is capable of regulating gene transcription.

Src-family tyrosine kinases are enzymes capable of phosphorylating tyrosine residues of a target molecule. Typically, a src-family tyrosine kinase contains one or more binding domains and a kinase domain. A binding domain of a src-family tyrosine kinase is capable of binding to a target molecule and a kinase domain is capable of phosphorylating a target molecule bound to the kinase. Members of the src family of tyrosine kinases are



characterized by an N-terminal unique region followed by three regions that contain different degrees of homology among all the members of the family. These three regions are referred to as src homology region 1 (SH1), src homology region 2 (SH2) and src homology region 3 (SH3). Both the SH2 and SH3 domains are believed to have protein association functions important for the formation of signal transduction complexes. The amino acid sequence of an N-terminal unique region, varies between each src-family tyrosine kinase. An N-terminal unique region can be at least about the first 40 amino acid residues of the N-terminal of a src-family tyrosine kinase.

Syk-family kinases are enzymes capable of phosphorylating tyrosine residues of a target molecule. Typically, a syk-family kinase contains one or more binding domains and a kinase domain. A binding domain of a syk-family tyrosine kinase is capable of binding to a target molecule and a kinase domain is capable of phosphorylating a target molecule bound to the kinase. Members of the syk-family of tyrosine kinases are characterized by two SH2 domains for protein association function and a tyrosine kinase domain.

A primary target molecule is capable of further extending a signal transduction pathway by modifying a second messenger molecule. Primary target molecules can include, but are not limited to, phosphatidylinositol 3-kinase (PI-3K), P21<sup>ras</sup>GAPase-activating protein and associated P190 and P62 protein, phospholipases such as PLC $\gamma$ 1 and PLC $\gamma$ 2, MAP kinase, Shc and VAV. A primary target molecule is capable of producing second messenger molecule which is capable of further amplifying a transduced signal. Second messenger molecules include, but are not limited to diacylglycerol and inositol 1,4,5-triphosphate (IP3). Second messenger molecules are capable of initiating physiological events which can lead to alterations in gene transcription. For example, production of IP3 can result in release of intracellular calcium, which can then lead to activation of calmodulin kinase II, which can then lead to serine phosphorylation of a DNA binding protein referred to as ets-1 proto-onco-protein. Diacylglycerol is capable of activating the signal transduction protein, protein kinase C which affects the activity of the AP1 DNA binding protein complex. Signal transduction pathways can lead to transcriptional activation of genes such as c-fos, egr-1, and c-myc.

Shc can be thought of as an adaptor molecule. An adaptor molecule comprises a protein that enables two other proteins to form a complex (e.g., a three molecule complex). Shc protein enables a complex to form which includes Grb2 and SOS. Shc comprises an SH2 domain that is capable of associating with the SH2 domain of Grb2.

Molecules of a signal transduction pathway can associate with one another using recognition sequences. Recognition sequences enable specific binding between two molecules. Recognition sequences can vary depending upon the structure of the molecules that are associating with one another. A molecule can have one or more recognition

sequences, and as such can associate with one or more different molecules.

Signal transduction pathways for MIRR complexes are capable of regulating the biological functions of a cell. Such functions can include, but are not limited to the ability of a cell to grow, to differentiate and to secrete cellular products. MIRR-induced signal transduction pathways can regulate the biological functions of specific types of cells involved in particular responses by an animal, such as immune responses, inflammatory responses and allergic responses. Cells involved in an immune response can include, for example, B cells, T cells, macrophages, dendritic cells, natural killer cells and plasma cells. Cells involved in inflammatory responses can include, for example, basophils, mast cells, eosinophils, neutrophils and macrophages. Cells involved in allergic responses can include, for example mast cells, basophils, B cells, T cells and macrophages.

In exemplary embodiments of the subject assay, the detection signal is a second messengers, such as a phosphorylated src-like protein, includes reporter constructs which include transcriptional regulatory elements such as serum response element (SRE), 12-O-tetradecanoyl-phorbol-13-acetate response element, cyclic AMP response element, c-fos promoter, or a CREB-responsive element.

## VII. Nuclear Receptors.

In another embodiment, the target receptor is a nuclear receptor. The nuclear receptors may be viewed as ligand-dependent transcription factors. These receptors provide a direct link between extracellular signals, mainly hormones, and transcriptional responses. Their transcriptional activation function is regulated by endogenous small molecules, such as steroid hormones, vitamin D, ecdysone, retinoic acids and thyroid hormones, which pass readily through the plasma membrane and bind their receptors inside the cell (Laudet and Adelman (1995) *Current Biology* 5:124). The majority of these receptors appear to contain three domains: a variable amino terminal domain, a highly conserved, DNA-binding domain and a moderately conserved, carboxyl-terminal ligand-binding domain (Power et al. (1993) *Curr. Opin. Cell Biol.* 5:499-504). Examples include the estrogen, progesterone, androgen, thyroid hormone and mineralocorticoid receptors. In addition to the known steroid receptors, at least 40 orphan members of this superfamily have been identified. (Laudet et al. (1992) *EMBO J.* 11:1003-1013). There are at least four groups of orphan nuclear receptors represented by NGF1, FTZ-F1, Rev-erbs, and RARs, which are by evolutionary standards, only distantly related to each other (Laudet et al. supra). While the steroid hormone receptors bind exclusively as homodimers to a palindrome of their hormone responsive element other nuclear receptors bind as heterodimers. Interestingly, some orphan receptors bind as monomers to similar response elements and require for their function a specific motif that is rich in basic amino-acid residues and is located carboxy-terminal to the DNA-binding domain (Laudet and Adelman supra.)

In preferred embodiments, the subject assay is derived to utilize a hormone-dependent reporter construct for selection. For instance, glucocorticoid response elements (GREs) and thyroid receptor enhancer-like DNA sequences (TREs) can be used to drive expression of reporter construct in response to hormone binding to hormone receptors. GREs are enhancer-like DNA sequences that confer glucocorticoid responsiveness via interaction with the glucocorticoid receptor. See Payvar, et al. (1983) *Cell* 35:381 and Schiedereit et al. (1983) *Nature* 304:749. TREs are similar to GREs except that they confer thyroid hormone responsiveness via interaction with thyroid hormone receptor. Turning now to the interaction of hormones and receptors, it is known that a steroid or thyroid hormone enters cells by facilitated diffusion and binds to its specific receptor protein, initiating an allosteric alteration of the protein. As a result of this alteration, the hormone/receptor complex is capable of binding to certain specific sites on transcriptional regulatory sequence with high affinity.

It is also known that many of the primary effects of steroid and thyroid hormones involve increased transcription of a subset of genes in specific cell types. Moreover, there is evidence that activation of transcription (and, consequently, increased expression) of genes which are responsive to steroid and thyroid hormones (through interaction of chromatin with hormone receptor/hormone complex) is effected through binding of the complex to enhancers associated with the genes.

In any case, a number of steroid hormone and thyroid hormone responsive transcriptional control units, some of which have been shown to include enhancers, have been identified. These include the mouse mammary tumor virus 5'-long terminal repeat (MTV-LTR), responsive to glucocorticoid, aldosterone and androgen hormones; the transcriptional control units for mammalian growth hormone genes, responsive to glucocorticoids, estrogens, and thyroid hormones; the transcriptional control units for mammalian prolactin genes and progesterone receptor genes, responsive to estrogens; the transcriptional control units for avian ovalbumin genes, responsive to progesterones; mammalian metallothionein gene transcriptional control units, responsive to glucocorticoids; and mammalian hepatic alpha 2u-globulin gene transcriptional control units, responsive to androgens, estrogens, thyroid hormones and glucocorticoids. Such steroid hormone and thyroid hormone responsive transcriptional control units can be used to generate reporter constructs which are sensitive to agonists and antagonists of the steroid hormone and/or thyroid hormone receptors. See, for example, U.S. Patents 5,298,429 and 5,071,773, both to Evans, et. al. Moreover, the art describes the functional expression of such receptors in yeast. See also for example, Caplan et al. (1995) *J Biol Chem* 270:5251-7; and Banihmad et al. (1995) *Mol Endocrinol* 9: 34-43.

#### VIII. Receptor tyrosine kinases.

In still another embodiment, the target receptor is a receptor tyrosine kinase. The

receptor tyrosine kinases can be divided into five subgroups on the basis of structural similarities in their extracellular domains and the organization of the tyrosine kinase catalytic region in their cytoplasmic domains. Sub-groups I (epidermal growth factor (EGF) receptor-like), II (insulin receptor-like) and the eph/cek family contain cysteine-rich sequences (Hirai et al., (1987) *Science* 238:1717-1720 and Hunter, (1990) *Mol. Cell. Biol.* 10:6316-6324). The functional domains of the kinase region of these three classes of receptor tyrosine kinases are encoded as a contiguous sequence (Hanks et al. (1988) *Science* 241:42-52). Subgroups III (platelet-derived growth factor (PDGF) receptor-like) and IV (the fibro-blast growth factor (FGF) receptors) are characterized as having immunoglobulin (Ig)-like folds in their extracellular domains, as well as having their kinase domains divided in two parts by a variable stretch of unrelated amino acids (Yanden and Ullrich (1988) *supra* and Hanks et al. (1988) *supra*).

The family with by far the largest number of known members is the EPH family. Since the description of the prototype, the EPH receptor (Hirai et al. (1987) *Science* 238:1717-1720), sequences have been reported for at least ten members of this family, not counting apparently orthologous receptors found in more than one species. Additional partial sequences, and the rate at which new members are still being reported, suggest the family is even larger (Maissonpierre et al. (1993) *Oncogene* 8:3277-3288; Andres et al. (1994) *Oncogene* 9:1461-1467; Henkemeyer et al. (1994) *Oncogene* 9:1001-1014; Ruiz et al. (1994) *Mech. Dev.* 46:87-100; Xu et al. (1994) *Development* 120:287-299; Zhou et al. (1994) *J. Neurosci. Res.* 37:129-143; and references in Tuzi and Gullick (1994) *Br. J. Cancer* 69:417-421). Remarkably, despite the large number of members in the EPH family, all of these molecules were identified as orphan receptors without known ligands.

The expression patterns determined for some of the EPH family receptors have implied important roles for these molecules in early vertebrate development. In particular, the timing and pattern of expression of *sek*, *mek4* and some of the other receptors during the phase of gastrulation and early organogenesis has suggested functions for these receptors in the important cellular interactions involved in patterning the embryo at this stage (Gilliardi-Hebensreit et al. (1992) *Oncogene* 7:2499-2506; Nieto et al. (1992) *Development* 116:1137-1150; Henkemeyer et al., *supra*; Ruiz et al., *supra*; and Xu et al., *supra*). *Sek*, for example, shows a notable early expression in the two areas of the mouse embryo that show obvious segmentation, namely the somites in the mesoderm and the rhombomeres of the hindbrain; hence the name *sek*, for segmentally expressed kinase (Gilliardi-Hebensreit et al., *supra*; Nieto et al., *supra*). As in *Drosophila*, these segmental structures of the mammalian embryo are implicated as important elements in establishing the body plan. The observation that *Sek* expression precedes the appearance of morphological segmentation suggests a role for *sek* in forming these segmental structures, or in determining segment-specific cell properties such as lineage compartmentation (Nieto et al., *supra*). Moreover, EPH receptors have been

implied, by their pattern of expression, in the development and maintenance of nearly every tissue in the embryonic and adult body. For instance, EPH receptors have been detected throughout the nervous system, the testes, the cartilaginous model of the skeleton, tooth primordia, the infundibular component of the pituitary, various epithelia tissues, lung, pancreas, liver and kidney tissues. Observations such as this have been indicative of important and unique roles for EPH family kinases in development and physiology, but further progress in understanding their action has been severely limited by the lack of information on their ligands.

As used herein, the terms "EPH receptor" or "EPH-type receptor" refer to a class of receptor tyrosine kinases, comprising at least eleven paralogous genes, though many more orthologs exist within this class, e.g. homologs from different species. EPH receptors, in general, are a discrete group of receptors related by homology and easily recognizable, e.g., they are typically characterized by an extracellular domain containing a characteristic spacing of cysteine residues near the N-terminus and two fibronectin type III repeats (Hirai et al. (1987) *Science* 238:1717-1720; Lindberg et al. (1990) *Mol Cell Biol* 10:6316-6324; Chan et al. (1991) *Oncogene* 6:1057-1061; Maisonnepierre et al. (1993) *Oncogene* 8:3277-3288; Andres et al. (1994) *Oncogene* 9:1461-1467; Henkemeyer et al. (1994) *Oncogene* 9:1001-1014; Ruiz et al. (1994) *Mech Dev* 46:87-100; Xu et al. (1994) *Development* 120:287-299; Zhou et al. (1994) *J Neurosci Res* 37:129-143; and references in Tuzi and Guillek (1994) *Br J Cancer* 69:417-421). Exemplary EPH receptors include the *eph*, *elk*, *eck*, *sek*, *mek4*, *hek*, *hek2*, *eek*, *erk*, *tyro1*, *tyro4*, *tyro5*, *tyro6*, *tyro11*, *cek4*, *cek5*, *cek6*, *cek7*, *cek8*, *cek9*, *cek10*, *bsk*, *rik1*, *rik2*, *rik3*, *myk1*, *myk2*, *ehk1*, *ehk2*, *pagliaccio*, *hik*, *erk* and *nuk* receptors. The term "EPH receptor" refers to the membrane form of the receptor protein, as well as soluble extracellular fragments which retain the ability to bind the ligand of the present invention.

In exemplary embodiments, the detection signal is provided by detecting phosphorylation of intracellular proteins, e.g., MEKKs, MEKs, or Map kinases, or by the use of reporter constructs which include transcriptional regulatory elements responsive to c-fos and/or c-jun. Described *infra*.

## IX. G Protein-Coupled Receptors.

One family of signal transduction cascades found in eukaryotic cells utilizes heterotrimeric "G proteins." Many different G proteins are known to interact with receptors. G protein signaling systems include three components: the receptor itself, a GTP-binding protein (G protein), and an intracellular target protein.

The cell membrane acts as a switchboard. Messages arriving through different receptors can produce a single effect if the receptors act on the same type of G protein. On the other hand, signals activating a single receptor can produce more than one effect if the

receptor acts on different kinds of G proteins, or if the G proteins can act on different effectors.

In their resting state, the G proteins, which consist of alpha ( $\alpha$ ), beta ( $\beta$ ) and gamma ( $\gamma$ ) subunits, are complexed with the nucleotide guanosine diphosphate (GDP) and are in contact with receptors. When a hormone or other first messenger binds to receptor, the receptor changes conformation and this alters its interaction with the G protein. This spurs the  $\alpha$  subunit to release GDP, and the more abundant nucleotide guanosine triphosphate (GTP), replaces it, activating the G protein. The G protein then dissociates to separate the  $\alpha$  subunit from the still complexed beta and gamma subunits. Either the G $\alpha$  subunit, or the G $\beta\gamma$  complex, depending on the pathway, interacts with an effector. The effector (which is often an enzyme) in turn converts an inactive precursor molecule into an active "second messenger," which may diffuse through the cytoplasm, triggering a metabolic cascade. After a few seconds, the G $\alpha$  converts the GTP to GDP, thereby inactivating itself. The inactivated G $\alpha$  may then reassociate with the G $\beta\gamma$  complex.

Hundreds, if not thousands, of receptors convey messages through heterotrimeric G proteins, of which at least 17 distinct forms have been isolated. Although the greatest variability has been seen in the  $\alpha$  subunit, several different  $\beta$  and  $\gamma$  structures have been reported. There are, additionally, several different G protein-dependent effectors. Most G protein-coupled receptors are comprised of a single protein chain that is threaded through the plasma membrane seven times. Such receptors are often referred to as seven-transmembrane receptors (STRs). More than a hundred different STRs have been found, including many distinct receptors that bind the same ligand, and there are likely many more STRs awaiting discovery.

In addition, STRs have been identified for which the natural ligands are unknown; these receptors are termed "orphan" G protein-coupled receptors, as described above. Examples include receptors cloned by Neote et al. (1993) *Cell* 72, 415; Koubas et al. *FEBS Lett.* (1993) 321, 173; Birkenbach et al. (1993) *J. Virol.* 67, 2209.

The "exogenous receptors" of the present invention may be any G protein-coupled receptor which is exogenous to the cell which is to be genetically engineered for the purpose of the present invention. This receptor may be a plant or animal cell receptor. Screening for binding to plant cell receptors may be useful in the development of, e.g., herbicides. In the case of an animal receptor, it may be of invertebrate or vertebrate origin. If an invertebrate receptor, an insect receptor is preferred, and would facilitate development of insecticides. The receptor may also be a vertebrate, more preferably a mammalian, still more preferably a human, receptor. The exogenous receptor is also preferably a seven transmembrane segment receptor.

Known ligands for G protein coupled receptors include: purines and nucleotides, such

as adenosine, cAMP, ATP, UTP, ADP, melatonin and the like; biogenic amines (and related natural ligands), such as 5-hydroxytryptamine, acetylcholine, dopamine, adrenaline, adrenaline, adrenaline, histamine, noradrenaline, noradrenaline, noradrenaline, tyramine/octopamine and other related compounds; peptides such as adrenocorticotrophic hormone (acth), melanocyte stimulating hormone (msh), melanocortins, neurotensin (nt), bombesin and related peptides, endothelins, cholecystokinin, gastrin, neuropeptide y (npy), thyrotropin releasing factor (trf), bradykinin, angiotensin ii, beta-endorphin, csa anaphalatoxin, calcitonin, chemokines (also called intercrines), corticotrophic releasing factor (crf), dynorphin, endorphin, tmlp and other formylated peptides, follitropin (fsh), fungal mating pheromones, galanin, gastric inhibitory polypeptide receptor (gip), glucagon-like peptides (glps), glucagon, gonadotropin releasing hormone (gnrh), growth hormone releasing hormone(ghrh), insect diuretic hormone, interleukin-8, leutropin (lh/hcg), met-enkephalin, opioid peptides, oxytocin, parathyroid hormone (pth) and pthrp, pituitary adenyl cyclase activating peptide (pacap), secretin, somatostatin, thrombin, thyrotropin (tsh), vasoactive intestinal peptide (vip), vasopressin, vasotocin; eicosanoids such as ip-prostaglyclin, pg-invertibrate 11-cis retinal and other related compounds; lipids and lipid-based compounds such as cannabinoids, anandamide, lysophosphatidic acid, platelet activating factor, leukotenes and the like; excitatory amino acids and ions such as calcium ions and glutamate.

Suitable examples of G-protein coupled receptors include, but are not limited to, dopaminergic, muscarinic cholinergic, a-adrenergic, b-adrenergic, opioid (including delta and mu), cannabinoid, serotoninergic, and GABAergic receptors. Preferred receptors include the 5HT family of receptors, dopamine receptors, C5a receptor and FPR1-1 receptor, cyclo-histidyl-proline-diketopiperazine receptors, melanocyte stimulating hormone release inhibiting factor receptor, and receptors for neurotensin, thyrotropin releasing hormone, calcitonin, cholecystokinin-A, neurokinin-2, histamine-3, cannabinoic, melanocortin, or adrenomedullin, neuropeptide-Y1 or galanin. Other suitable receptors are listed in the art. The term "receptor," as used herein, encompasses both naturally occurring and mutant receptors. Many of these G protein-coupled receptors, like the yeast a- and a-factor receptors, contain seven hydrophobic amino acid-rich regions which are assumed to lie within the plasma membrane. Specific human G protein-coupled STRs for which genes have been isolated and for which expression vectors could be constructed include those listed herein and others known in the art. Thus, the gene would be operably linked to a promoter functional in the cell to be engineered and to a signal sequence that also functions in the cell. For example in the case of yeast, suitable promoters include Ste2, Ste3 and gal10. Suitable signal sequences include those of Ste2, Ste3 and of other genes which encode proteins secreted by

yeast cells. Preferably, when a yeast cell is used, the codons of the gene would be optimized for expression in yeast. See Hoekema et al., (1987) *Mol. Cell. Biol.*, 7:2914-24; Sharp, et al., (1986) 14:5125-43.

The homology of STRs is discussed in Dohman et al., *Ann. Rev. Biochem.*, (1991) 60:653-88. When STRs are compared, a distinct spatial pattern of homology is discernible. The transmembrane domains are often the most similar, whereas the N- and C-terminal regions, and the cytoplasmic loop connecting transmembrane segments V and VI are more divergent.

The functional significance of different STR regions has been studied by introducing point mutations (both substitutions and deletions) and by constructing chimeras of different but related STRs. Synthetic peptides corresponding to individual segments have also been tested for activity. Affinity labeling has been used to identify ligand binding sites.

It is conceivable that a foreign receptor which is expressed in yeast will functionally integrate into the yeast membrane, and there interact with the endogenous yeast G protein. More likely, either the receptor will need to be modified (e.g., by replacing its V-VI loop with that of the yeast STE2 or STE3 receptor), or a compatible G protein should be provided. If the wild-type exogenous G protein-coupled receptor cannot be made functional in yeast, it may be mutated for this purpose. A comparison would be made of the amino acid sequences of the exogenous receptor and of the yeast receptors, and regions of high and low homology identified. Trial mutations would then be made to distinguish regions involved in ligand or G protein binding, from those necessary for functional integration in the membrane. The exogenous receptor would then be mutated in the latter region to more closely resemble the yeast receptor, until functional integration was achieved. If this were insufficient to achieve functionality, mutations would next be made in the regions involved in G protein binding. Mutations would be made in regions involved in ligand binding only as a last resort, and then an effort would be made to preserve ligand binding by making conservative substitutions whenever possible.

Preferably, the yeast genome is modified so that it is unable to produce the yeast receptors which are homologous to the exogenous receptors in functional form. Otherwise, a positive assay score might reflect the ability of a peptide to activate the endogenous G protein-coupled receptor, and not the receptor of interest.

#### A. Chemottractant receptors

The N-formyl peptide receptor is a classic example of a calcium mobilizing G protein-coupled receptor expressed by neutrophils and other phagocytic cells of the mammalian immune system (Snyderman et al. (1988) *In Inflammation: Basic Principles and*



*Clinical Correlates*, pp. 309-323). N-formyl peptides of bacterial origin bind to the receptor and engage a complex activation program that results in directed cell movement, release of inflammatory granule contents, and activation of a latent NADPH oxidase which is important for the production of metabolites of molecular oxygen. This pathway initiated by receptor-ligand interaction is critical in host protection from pyogenic infections. Similar signal transduction occurs in response to the inflammatory peptides C5a and IL-8.

Two other formyl peptide receptor like (FPR) genes have been cloned based on their ability to hybridize to a fragment of the NPPR cDNA coding sequence. These have been named FPR1 (Murphy et al. (1992) *J. Biol. Chem.* 267:7637-7643) and FPR2 (Ye et al. (1992) *Biochem Biophys Res. Comm.* 184:582-589). FPR2 was found to mediate calcium mobilization in mouse fibroblasts transfected with the gene and exposed to formyl peptide. In contrast, although FPR1 was found to be 69% identical in amino acid sequence to NPPR, it did not bind prototype N-formyl peptides ligands when expressed in heterologous cell types. This led to the hypothesis of the existence of an as yet unidentified ligand for the FPR1 orphan receptor (Murphy et al. *supra*).

Using the technology described herein a ligand has been cloned for these orphan receptors.

## B. G proteins

In the case of an exogenous G-protein coupled receptor, the yeast cell must be able to produce a G protein which is activated by the exogenous receptor, and which can in turn activate the yeast effector(s). The art suggests that the endogenous yeast G subunit (e.g., GPA) will be often be sufficiently homologous to the "cognate" G subunit which is natively associated with the exogenous receptor for coupling to occur. More likely, it will be necessary to genetically engineer the yeast cell to produce a foreign G subunit which can properly interact with the exogenous receptor. For example, the G subunit of the yeast G protein may be replaced by the G subunit natively associated with the exogenous receptor.

Dietzel and Kujan, (1987) *Cell*, 50:1001) demonstrated that rat Gas functionally coupled to the yeast Gby complex. However, rat Ga12 complemented only when substantially overexpressed, while Ga0 did not complement at all. Kang, et al., *Mol. Cell. Biol.*, (1990)10:2582). Consequently, with some foreign G subunits, it is not feasible to simply replace the yeast Gα.

If the exogenous G protein coupled receptor is not adequately coupled to yeast Gby by the Gα subunit natively associated with the receptor, the Gα subunit may be modified to improve coupling. These modifications often will take the form of mutations which increase the resemblance of the Gα subunit to the yeast Gα while decreasing its resemblance to the

receptor-associated Gα. For example, a residue may be changed so as to become identical to the corresponding yeast Gα residue, or to at least belong to the same exchange group of that residue. After modification, the modified Gα subunit might or might not be "substantially homologous" to the foreign and/or the yeast Gα subunit.

The modifications are preferably concentrated in regions of the Gα which are likely to be involved in Gβγ binding. In some embodiments, the modifications will take the form of replacing one or more segments of the receptor-associated Gα with the corresponding yeast Gα segment(s), thereby forming a chimeric Gα subunit. (For the purpose of the appended claims, the term "segment" refers to three or more consecutive amino acids.) In other embodiments, point mutations may be sufficient.

This chimeric Gα subunit will interact with the exogenous receptor and the yeast Gβγ complex, thereby permitting signal transduction. While use of the endogenous yeast Gβγ is preferred, if a foreign or chimeric Gβγ is capable of transducing the signal to the yeast effector, it may be used instead.

### C. Gα Structure

Some aspects of Gα structure are relevant to the design of modified Gα subunits. The amino terminal 66 residues of GPA1 are aligned with the cognate domains of human Gαs, Gαi2, Gαi3, Gαi6 and *transducin*. In the GPA4/Gα hybrids, the amino terminal 41 residues (derived from GPA1) are identical, end with the sequence-LKQQRDKNE- and are underlined for emphasis. All residues following the glutamate (E) residue at position 41 are contributed by the human Gα subunits, including the consensus nucleotide binding motif -GXGxxG-. Periods in the sequences indicate gaps that have been introduced to maximize alignments in this region. Codon bias is mammalian. For alignments of the entire coding regions of GPA1 with Gαs, Gαi, and GαO, Gαq and Gαz, see Dietzel and Kuryan (1987, *Cell* 50:573) and Lambright, et al. (1994, *Nature* 369:621-628). Additional sequence information is provided by Mattern, et al. (1986, *FEBS Lett* 206:36-41), Bray, et al. (1986, *Proc. Natl. Acad. Sci USA* 83:8893-8897) and Bray, et al. (1987, *Proc Natl. Acad Sci USA* 84:5115-5119).

The gene encoding a G protein homolog of *S. cerevisiae* was cloned independently by Dietzel and Kuryan (*supra*) (SCG1) and by Nakafuku, et al. (1987 *Proc Natl Acad Sci* 84:2140-2144) (GPA1). Sequence analysis revealed a high degree of homology between the protein encoded by this gene and mammalian Gα. GPA1 encodes a protein of 472 amino acids, as compared with approximately 340-350 a.a. for most mammalian Gα subunits in four described families, Gαs, Gαi, Gαq and Gα12/13. Nevertheless, GPA1 shares overall sequence and structural homology with all Gα proteins identified to date. The highest overall homology in GPA1 is to the Gαi family (48% identity, or 65% with conservative

substitutions) and the lowest is to GQS (33% identity, or 51% with conservative substitutions) (Nakafuku, et al., *supra*).

The regions of high sequence homology among Gα subunits are dispersed throughout their primary sequences, with the regions sharing the highest degree of homology mapping to sequence that comprises the guanine nucleotide binding/GTPase domain. This domain is structurally similar to the αβ fold of Ras proteins and the protein synthesis elongation factor EF-Tu. This highly conserved guanine nucleotide-binding domain consists of a six-stranded β sheet surrounded by a set of five α-helices. It is within these β sheets and α helices that the highest degree of conservation is observed among all Gα proteins, including GAP. The least sequence and structural homology is found in the intervening loops between the β sheets and α helices that define the core GTPase domain. There are a total of four "intervening loops" or "inserts" present in all Gα subunits. In the crystal structures reported to date for the GDP- and GTPγS-liganded forms of bovine rod transducin (Noel, et al. (1993) *Nature* 366:654-663); (Lambright, et al. (1994) *Nature* 369:621-628), the loop residues are found to be outside the core GTPase structure. Functional roles for these loop structures have been established in only a few instances. A direct role in coupling to phosphodiesterase-γ has been demonstrated for residues within inserts 3 and 4 of Gat (Ratick, et al. (1992) *Science* 256:1031-1033); (Artemyev, et al. (1992) *J. Biol. Chem.* 267:25067-25072), while a "GAP-like" activity has been ascribed to the largely α-helical insert 1 domain of Gαs (Markby, et al. (1993) *Science* 262:1805-1901).

While the amino- and carboxy-termini of Gα subunits do not share striking homology either at the primary, secondary, or tertiary levels, there are several generalizations that can be made about them. First, the amino termini of Gα subunits have been implicated in the association of Gα with Gβγ complexes and in membrane association via N-terminal myristoylation. In addition, the carboxy-termini have been implicated in the association of Gαβγ heterotrimeric complexes with G protein-coupled receptors (Sullivan, et al. (1987) *Nature* 330:758-760); West, et al. (1985) *J. Biol. Chem.* 260:14428-14430); (Conklin, et al. (1993) *Nature* 363:274-276). Data in support of these generalizations about the function of the N-terminus derive from several sources, including both biochemical and genetic studies.

As indicated above, there is little if any sequence homology shared among the amino termini of Gα subunits. The amino terminal domains of Gα subunits that precede the first β-sheet (containing the sequence motif -LLLLGAGESG<sup>(see 10 No. 1)</sup>; see Noel, et al. (*supra*) for the numbering of the structural elements of Gα subunits) vary in length from 41 amino acids (GPA1) to 31 amino acids (Gat). Most Gα subunits share the consensus sequence for the addition of myristic acid at their amino termini (MGxxxS-), although not all Gα subunits that contain this motif have myristic acid covalently associated with the glycine at position 2 (Speigel, et al. (1991) *TIBS* 16:338-344). The role of this post-translational modification has

been inferred from studies in which the activity of mutant Gα subunits from which the consensus sequence for myristoylation has been added or deleted has been assayed (Mumby et al. (1990) *Proc. Natl. Acad. Sci. USA* 87: 728-732; Linder, et al. (1991) *J. Biol. Chem.* 266:4654-4659); Gallego, et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:9695-9699). These studies suggest two roles for N-terminal myristoylation. First, the presence of amino-terminal myristic acid has in some cases been shown to be required for association of Gα subunits with the membrane, and second, this modification has been demonstrated to play a role in modulating the association of Gα subunits with Gβγ complexes. The role of myristoylation of the GPA1 gene products, at present, unknown.

In other biochemical studies aimed at examining the role of the amino-terminus of Gα in driving the association between Gα and Gβγ subunits, proteolytically or genetically truncated versions of Gα subunits were assayed for their ability to associate with Gβγ complexes, bind guanine nucleotides and/or to activate effector molecules. In all cases, Gα subunits with truncated amino termini were deficient in all three functions (Graf, et al. (1992) *J. Biol. Chem.* 267:24307-24314); (Journot, et al. (1990) *J. Biol. Chem.* 265:9009-9015); and (Neer, et al. (1988) *J. Biol. Chem.* 263:8996-9000). Slepak, et al. (1993, *J. Biol. Chem.* 268:1414-1423) reported a mutational analysis of the N-terminal 56 a.a. of mammalian Gαo expressed in *Escherichia coli*. Molecules with an apparent reduced ability to interact with exogenously added mammalian Gβγ were identified in the mutant library. As the authors pointed out, however, the assay used to screen the mutants the extent of ADP-ribosylation of the mutant Gα by pertussis toxin was not a completely satisfactory probe of interactions between Gα and Gβγ. Mutations identified as inhibiting the interaction of the subunits, using this assay, may still permit the complexing of Gα and Gβγ while sterically hindering the ribosylation of Gα by toxin. Genetic studies examined the role of amino-terminal determinants of Gα in heterotrimer subunit association have been carried out in both yeast systems using GPA1-mammalian Gα hybrids (Kang, et al. (1990) *Mol. Cell. Biol.* 10:2582-2590) and in mammalian systems using Gαi/Gαs hybrids (Russell and Johnson (1993) *Mol. Pharmacol.* 44:255-263). In the former studies, gene fusions, composed of yeast GPA1 and mammalian Gα sequences were constructed by Kang, et al. (*supra*) and assayed for their ability to complement a gpa1 null phenotype (i.e., constitutive activation of the pheromone response pathway) in *S. cerevisiae*. Kang, et al. demonstrated that wild type mammalian Gαs, Gαi but not Gαo proteins are competent to associate with yeast Gα and suppress the gpa1 null phenotype, but only when overexpressed. Fusion proteins containing the amino-terminal 330 residues of GPA1 sequence linked to 160, 143, or 142 residues of the mammalian Gαs, Gαi and Gαo carboxyl-terminal regions, respectively, also coupled to the yeast mating response pathway when overexpressed on high copy plasmids with strong inducible (CUP) or constitutive (PGK) promoters. All three of these hybrid molecules were able to complement the gpa1 null mutation in a growth arrest assay, and were additionally

able to inhibit  $\alpha$  factor responsiveness and mating in tester strains. These last two observations argue that hybrid yeast-mammalian G $\alpha$  subunits are capable of interacting directly with yeast G $\beta\gamma$ , thereby disrupting the normal function of the yeast heterotrimer. Fusions containing the amino terminal domain of G $\alpha_s$ , G $\alpha_i$  or G $\alpha_o$ , however, did not complement the gpa1 null phenotype, indicating a requirement for determinants in the amino terminal 330 amino acid residues of GPA1 for association and sequestration of yeast G $\beta\gamma$  complexes. Taken together, these data suggest that determinants in the amino terminal region of G $\alpha$  subunits determine not only the ability to associate with G $\beta\gamma$  subunits in general, but also with specific G $\beta\gamma$  subunits in a species-restricted manner.

Hybrid G $\alpha_i$ /G $\alpha_s$  subunits have been assayed in mammalian expression systems (Russell and Johnson *supra*). In these studies, a large number of chimeric G $\alpha$  subunits were assayed for an ability to activate adenylate cyclase, and therefore, for an ability to interact with G $\beta\gamma$ (i.e., coupling of G $\alpha$  to G $\beta\gamma$ =inactive cyclase; uncoupling of G $\alpha$  from G $\beta\gamma$ =active cyclase). From these studies a complex picture emerged in which determinants in the region between residues 25 and 96 of the hybrids were found to determine the state of activation of these alleles as reflected in their rates of guanine nucleotide exchange and GTP hydrolysis and the extent to which they activated adenylate cyclase *in vivo*. These data could be interpreted to support the hypothesis that structural elements in the region between the amino terminal methionine and the ~1 sheet identified in the crystal structure of G $\alpha_t$  (see Noel, et al. *supra* and Lambright, et al. *supra*) are involved in determining the state of activity of the heterotrimer by (1) driving association/dissociation between G $\alpha$  and G $\beta\gamma$  subunits; (2) driving GDP/GTP exchange. While there is no direct evidence provided by these studies to support the idea that residues in this region of G $\alpha$  and residues in G $\beta\gamma$  subunits contact one another, the data nonetheless provide a positive indication for the construction of hybrid G $\alpha$  subunits that retain function. There is, however, a negative indicator that derives from this work in that some hybrid constructs resulted in constitutive activation of the chimeric proteins (i.e., a loss of receptor-dependent stimulation of G $\beta\gamma$  dissociation and effector activation).

#### D. Construction of chimeric G $\alpha$ subunits.

In designing G $\alpha$  subunits capable of transmitting, in yeast, signals originating at mammalian G protein-coupled receptors, two general desiderata were recognized. First, the subunits should retain as much of the sequence of the native mammalian proteins as possible. Second, the level of expression for the heterologous components should approach, as closely as possible, the level of their endogenous counterparts. The results described by King, et al. (1990, Science 250:121-123) for expression of the human  $\beta_2$ -adrenergic receptor and G $\alpha_s$  in yeast, taken together with negative results obtained by Kang, et al. (*supra*) with full-length

mammalian Gα subunits other than Gαs, led us to the following preferences for the development of yeast strains in which mammalian G protein-coupled receptors could be linked to the pheromone response pathway.

1. Mammalian Gα subunits will be expressed using the native sequence of each subunit or, alternatively, as minimal gene fusions with sequences from the amino-terminus of GPA1 replacing the homologous residues from the mammalian Gα subunits.
2. Mammalian Gα subunits will be expressed from the GPA1 promoter either on low copy plasmids or after integration into the yeast genome as a single copy gene.
3. Endogenous Gβγ subunits will be provided by the yeast STE4 and STE18 loci.

#### *E. Site-Directed Mutagenesis versus Random Mutagenesis*

There are two general approaches to solving structure-function problems of the sort presented by attempts to define the determinants involved in mediating the association of the subunits that comprise the G protein heterotrimer. The first approach, discussed above with respect to hybrid constructs, is a rational one in which specific mutations or alterations are introduced into a molecule based upon the available experimental evidence. In a second approach, random mutagenesis techniques, coupled with selection or screening systems, are used to introduce large numbers of mutations into a molecule, and that collection of randomly mutated molecules is then subjected to a selection for the desired phenotype or a screen in which the desired phenotype can be observed against a background of undesirable phenotypes. With random mutagenesis one can mutagenize an entire molecule or one can proceed by cassette mutagenesis. In the former instance, the entire coding region of a molecule is mutagenized by one of several methods (chemical, PCR, doped oligonucleotide synthesis) and that collection of randomly mutated molecules is subjected to selection or screening procedures. Random mutagenesis can be applied in this way in cases where the molecule being studied is relatively small and there are powerful and stringent selections or screens available to discriminate between the different classes of mutant phenotypes that will inevitably arise. In the second approach, discrete regions of a protein, corresponding either to defined structural (i.e. α-helices, β-sheets, turns, surface loops) or functional determinants (e.g., catalytic clefts, binding determinants, transmembrane segments) are subjected to saturating or semi-random mutagenesis and these mutagenized cassettes are re-introduced into the context of the otherwise wild type allele. Cassette mutagenesis is most useful when there is experimental evidence available to suggest a particular function for a region of a molecule and there is a powerful selection and/or screening approach available to discriminate between interesting and uninteresting mutants. Cassette mutagenesis is also useful when the parent molecule is comparatively large and the desire is to map the functional domains of a molecule by mutagenizing the molecule in a step-wise fashion, i.e.

mutating one linear cassette of residues at a time and then assaying for function. The present invention contemplates applying random mutagenesis in order to further delineate the determinants involved in Gα-Gβγ association. Random mutagenesis may be accomplished by many means, including:

1. PCR mutagenesis, in which the error prone Taq polymerase is exploited to generate mutant alleles of Gα subunits, which are assayed directly in yeast for an ability to couple to yeast Gβγ.
2. Chemical mutagenesis, in which expression cassettes encoding Gα subunits are exposed to mutagens and the protein products of the mutant sequences are assayed directly in yeast for an ability to couple to yeast Gβγ.
3. Doped synthesis of oligonucleotides encoding portions of the Gα gene.
4. *In vivo* mutagenesis, in which random mutations are introduced into the coding region of Gα subunits by passage through a mutator strain of *E. coli*, XLI-Red (mutD5 mutS mutT) (Stratagene, Menasa, WI).

The random mutagenesis may be focused on regions suspected to be involved in Gα-Gβγ association as discussed in the next section. Random mutagenesis approaches are feasible for two reasons. First, in yeast one has the ability to construct stringent screens and facile selections (growth vs. death, transcription vs. lack of transcription) that are not readily available in mammalian systems. Second, when using yeast it is possible to screen efficiently through thousands of transformants rapidly. Cassette mutagenesis is immediately suggested by the observation (see *infra*) that the GPA41 hybrids couple to the pheromone response pathway. This relatively small region of Gα subunits represents a reasonable target for this type of mutagenesis. Another region that may be amenable to cassette mutagenesis is that defining the surface of the switch region of Gα subunits that is solvent-exposed in the crystal structures of Gαi and transducin. From the data described below, this surface may contain residues that are in direct contact with yeast Gβγ subunits, and may therefore be a reasonable target for mutagenesis.

#### *F. Rational Design of Chimeric Gα Subunits*

Several classes of rationally designed GPAI-mammalian Gα hybrid subunits have been tested for the ability to couple to yeast Gβγ. The first, and largest, class of hybrids are those that encode different lengths of the GPAI amino terminal domain in place of the homologous regions of the mammalian Gα subunits. This class of hybrid molecules includes GPABAMHI, GPA41, GPAID, and GPALW hybrids, described below. The rationale for constructing these hybrid Gα proteins is based on results, described above, that bear on the importance of the amino terminal residues of Gα in mediating interaction with Gβγ.

Preferably, the yeast Gα subunit is replaced by a chimeric Gα subunit in which a portion, e.g., at least about 20, more preferably at least about 40, amino acids, which is substantially homologous with the corresponding residues of the amino terminus of the yeast Gα, is fused to a sequence substantially homologous with the main body of a mammalian (or other exogenous) Gα. While 40 amino acids is the suggested starting point, shorter or longer portions may be tested to determine the minimum length required for coupling to yeast Gβγ and the maximum length compatible with retention of coupling to the exogenous receptor. It is presently believed that only the final 10 or 20 amino acids at the carboxy terminus of the Gα subunit are required for interaction with the receptor.

**GPA41 hybrids.** Kang et al. *supra*, described hybrid Gα subunits encoding the amino terminal 310 residues of GPAI fused to the carboxyl terminal 160, 143 and 142 residues, respectively, of Gαs, Gαi2, and Gαo. In all cases examined by Kang et al., the hybrid proteins were able to complement the growth arrest phenotype of *gpa1* strains. We have confirmed these findings and, in addition, have constructed and tested hybrids between GPAI and Gαi3, Gαq and Gα16. All hybrids of this type that have been tested functionally complement the growth arrest phenotype of *gpa1* strains.

GPA41 hybrids. The rationale for constructing a minimal hybrid encoding only 41 amino acids of GPAI relies upon the biochemical evidence for the role of the amino-terminus of Gα subunits discussed above, together with the following observation. Gβ and Gγ subunits are known to interact via α-helical domains at their respective amino-termini (Pronin, et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6220-6224); Garritsen, et al. (1993). The suggestion that the amino termini of Gα subunits may form an helical coil and that this helical coil may be involved in association of Gα with Gβγ (Masters et al (1986) *Protein Engineering* 1:47-54); Lupas et al. (1992) *FEBS Lett.* 314:105-108) leads to the hypothesis that the three subunits of the G-protein heterotrimer interact with one another reversibly through the winding and unwinding of their amino-terminal helical regions. A mechanism of this type has been suggested, as well, from an analysis of leucine zipper mutants of the GCN4 transcription factor (Harbury, et al. (1993) *Science* 262:1401-1407). The rationale for constructing hybrids like those described by Kang, et al. *supra*, that contain a majority of yeast sequence and only minimal mammalian sequence, derives from their ability to function in assays of coupling between Gα and Gβγ subunits. However, these chimeras had never been assayed for an ability to couple to both mammalian G protein-coupled receptors and yeast Gβγ subunits, and hence to reconstitute a hybrid signaling pathway in yeast. GPA41 hybrids that have been constructed and tested include Gαs, Gαi2, Gαi3, Gαq, Gαo, Gαop and Gα16. Hybrids of Gαs, Gαi2, Gαi3, and Gα16 functionally complement the growth arrest phenotype of *gpa1* strains, while GPA41 hybrids of Gαo, Gαop and Gαq do not. In addition to being tested in a growth arrest assay, these constructs have been assayed in



the more sensitive transcriptional assay for activation of a *fus1p-HIS3* gene. In both of these assays, the GPA41-Gas hybrid couples less well than the GPA41-I2, -I3, and -I6 hybrids, while the GPA41-O<sub>a</sub> and -O<sub>b</sub> hybrids do not function in either assay.

Several predictive algorithms indicate that the amino terminal domain up to the highly conserved sequence motif-LLLLGAGESG (the first L in this motif is residue 43 in GPA1) forms a helical structure with amphipathic character. Assuming that a heptahelical repeat unit, the following hybrids between GPA1 and Gas can be used to define the number of helical repeats in this motif necessary for hybrid function:

GPA1-7/Gas8-394  
GPA1-14/Gas15-394  
GPA1-21/Gas22-394  
GPA1-28/Gas29-394  
GPA1-35/Gas36-394  
GPA1-42/Gas43-394

In these hybrids, the prediction is that the structural repeat unit in the amino terminal domain up to the tetra-leucine motif is 7, and that swapping sequences in units of 7 will in effect amount to a swap of unit turns of the helical structure that comprises this domain.

A second group of "double crossover" hybrids of this class are those that are aligned on the first putative heptad repeat beginning with residue G11 in GPA1. In these hybrids, helical repeats are swapped from GPA1 into a Gas backbone one heptad repeat unit at a time.

GAs1-10/GPA11-17/GAs18-394

GAs1-17/GPA18-24/GAs25-394

GAs1-17/GPA25-31/GAs32-394

GAs7-17/GPA32-38/GAs39-394

The gap that is introduced between residues 9 and 10 in the Gas sequence is to preserve the alignment of the -LLLLGAGG-sequence motif. This class of hybrids can be complemented by cassette mutagenesis of each heptad repeat followed by screening of these collections of "heptad" libraries in standard coupling assays.

A third class of hybrids based on the prediction that the amino terminus forms a helical domain with a heptahelical repeat unit are those that effect the overall hydrophobic or hydrophilic character of the opposing sides of the predicted helical structure (See Lupas et al. *supra*). In this model, the  $\alpha$  and d positions of the heptad repeat abcdefg are found to be conserved hydrophobic residues that define one face of the helix, while the e and g positions

mammalian subunits, and therefore do not contain the GPA1 il region, while the GPALW hybrids contain the amino terminal 244 residues of GPA1 including the entire il region (as defined by sequence alignments). Hybrids of both GPAID and GPALW classes were constructed for Gαs, C-α12, Gαi3, Gαo, and Gα16; none of these hybrids complemented the gpa1 growth arrest phenotype.

Subsequent to the construction and testing of the GPAID and GPALW classes of hybrids, the crystal structures of transducin in both the GDP and GTPS-*liganded* form, and the crystal structure of several Gαi variants in the GTPS-*liganded* and GDP-*AlF<sub>4</sub>* forms were reported (Noel et al. *supra*; Lambright et al. *supra*; and Coleman et al. (1994) *Science* 265:1405-1412). The crystal-structures reveal that the il region defined by sequence alignment has a conserved structure that is comprised of six alpha helices in a rigid array, and that the junctions chosen for the construction of the GPAID and GPALW hybrids were not compatible with conservation of the structural features of the il region observed in the crystals. The junction chosen for the GPAID hybrids falls in the center of the long αA helix; chimerization of this helix in all likelihood destabilizes it and the protein structure in general. The same is true of the junction chosen for the GPALW hybrids in which the crossover point between GPA1 and the mammalian Gα subunit falls at the end of the short αC helix and therefore may distort it and destabilize the protein.

The failure of the GPAID and GPALW hybrids is predicted to be due to disruption of critical structural elements in the il region as discussed above. Based upon new alignments and the data presented in Noel et al (*supra*), Lambright et al (*supra*), and Coleman et al (*supra*), this problem can be averted with the ras-like core domain and the il helical domain are introduced outside of known structural elements like alpha-helices.

#### Hybrid A GαS1-67/GPA66-299/GαS203-394

This hybrid contains the entire il insert of GPA1 interposed into the GαS sequence.

#### Hybrid B GPA1-41/GαS4443-67/GPA66-299/GαS203-394

This hybrid contains the amino terminal 41 residues of GPA1 in place of the 42 amino terminal residues of GαS found in Hybrid A.

GαS Hybrids. There is evidence that the "switch region" encoded by residues 171-237 of Gα transducin (using the numbering of (Noel et al (*supra*) also plays a role in Gβγ coupling. First, the G226A mutation in GαS prevents the GTP-induced conformational change that occurs with exchange of GDP for GTP upon receptor activation by ligand. This residue maps to the highly conserved sequence -DVGGQ-<sup>Y</sup> present in all Gα subunits and is involved in

GTP hydrolysis. In both the Gat and Gcs crystal structures, this sequence motif resides in the loop that connects the  $\beta 3$  sheet and the  $\alpha 2$  helix in the guanine nucleotide binding core. In addition to blocking the conformational change that occurs upon GTP binding, this mutation also prevents dissociation of GTP-liganded Gcs from Gby. Second, crosslinking data reveals that a highly conserved cysteine residue in the  $\alpha 2$  helix (C215 in Gcs, C210 in Gat) can be crosslinked to the carboxy terminal region of Gb subunits. Finally, genetic evidence (Whiteaway et al. (1993) Mol Cell Biol. 14:3233-3239) identifies an important single residue in GPA1 (E307) in the  $\beta 2$  sheet of the core structure that may be in direct contact with Gby. A mutation in the GPA1 protein at this position suppresses the constitutive signalling phenotype of a variety of STE4 (Gb) dominant negative mutations that are also known to be defective in Gcs-Gby association (as assessed in two-hybrid assay in yeast as well as by more conventional genetic tests).

We have tested the hypothesis that there are switch region determinants involved in the association of Gcs with Gby by constructing a series of hybrid Gcs proteins encoding portions of GPA1 and Gcs in different combinations.

Two conclusions may be drawn. First, in the context of the amino terminus of Gcs, the GPA1 switch region suppresses coupling to yeast Gby (SGS), while in the context of the GPA1 amino terminus the GPA1 switch region stabilizes coupling with Gby (GPby-SGS). This suggests that these two regions of GPA1 collaborate to allow interactions between Gcs subunits and Gby subunits. This conclusion is somewhat mitigated by the observation that the GPA1-Gcs hybrid that does not contain the GPA1 switch region is able to complement the growth arrest phenotype of gpa1 strains. We have not to date noted a quantitative difference between the behavior of the GPA1-Gcs allele and the GPA1-SGS allele, but if this interaction is somewhat degenerate, then it may be difficult to quantitate this accurately. The second conclusion that can be drawn from these results is that there are other determinants involved in stabilizing the interaction of Gcs with Gby beyond these two regions as none of the GPA1/Gcs hybrid proteins couple as efficiently to yeast Gby as does native GPA1.

The role of the surface-exposed residues of this region may be crucial for effective coupling to yeast Gby, and can be incorporated into hybrid molecules as follows below.

Gcs-GPA-Switch Gcs 1-202/GPA298-350/Gcs 253-394

This hybrid encodes the entire switch region of GPA1 in the context of Gcs.

Gcs-GPA- $\alpha 2$  Gcs 1-226/GPA322-332/Gcs 238-394

This hybrid encodes the  $\alpha 2$  helix of GPA1 in the context of Gcs.

This hybrid encodes the 41 residue amino terminal domain of GPA1 and the α2 helix of GPA1 in the context of GαS.

Finally, the last class of hybrids that will be discussed here are those that alter the surface exposed residues of the β2 and β3 sheets of αS so that they resemble those of the GPA1 Q5 helix. These altered α2 helical domains have the following structure. (The positions of the altered residues correspond to GαS.)

L203K  
K211E  
D215G  
K216S  
D229S

These single mutations can be engineered into a GαS backbone singly and in pairwise combinations. In addition, they can be introduced in the context of both the full length GαS and the GPA41-GαS hybrid described previously. All are predicted to improve the coupling of Gα subunits to yeast Gβγ subunits by virtue of improved electrostatic and hydrophobic contacts between this region and the regions of Gβ defined by Whiteaway and coworkers (Whiteaway et al (*supra*) that define site(s) that interact with GPA1).

In summary, the identification of hybrid Gα subunits that couple to the yeast pheromone pathway has led to the following general observations. First, all GPABAMH1 hybrids associate with yeast Gβγ, therefore at a minimum these hybrids contain the determinants in GPA1 necessary for coupling to the pheromone response pathway. Second, the amino terminal 41 residues of GPA1 contain sufficient determinants to facilitate coupling of Gα hybrids to yeast Gβγ in some, but not all, instances, and that some Gα subunits contain regions outside of the first 41 residues that are sufficiently similar to those in GPA1 to facilitate interaction with GPA1 even in the absence of the amino terminal 41 residues of GPA1. Third, there are other determinants in the first 310 residues of GPA1 that are involved in coupling Gα subunits to yeast Gβγ subunits.

The various classes of hybrids noted above are not mutually exclusive. For example, a GPA1 containing GPA1-41 could also feature the L203K mutation.

While, for the sake of simplicity, we have described hybrids of yeast GPA1 and a mammalian Gαs, it will be appreciated that hybrids may be made of other yeast Gα subunits and/or other mammalian Gα subunits, notably mammalian Gαi subunits. Moreover, while the described hybrids are constructed from two parental proteins, hybrids of three or more parental proteins are also possible.

As shown in the Examples, chimeric Gα subunits have been especially useful in

*G. Expression of Gα*

Kang et al. *supra* reported that several classes of native mammalian G~ subunits were able to interact functionally with yeast α subunits when expression of Gα was driven from a constitutively active, strong promoter (PGK) or from a strong inducible promoter (CUP). These authors reported that rat Gαs, Gαi2 or Gαo expressed at high level coupled to yeast Gβγ. High level expression of mammalian Gα (i.e. non-stoichiometric with respect to yeast Gβγ) is not desirable for uses like those described in this application. Reconstruction of G protein-coupled receptor signal transduction in yeast requires the signalling component of the heterotrimeric complex (Gβγ) to be present stoichiometrically with Gα subunits. An excess of Gα subunits (as was required for coupling of mammalian Gαi2 and Gαo to yeast Gβγ in Kang et al.) would dampen the signal in systems where Gβγ subunits transduce the signal. An excess of Gα subunits raises the background level of signaling in the system to unacceptably high levels. Preferably, levels of Gα and Gβγ subunits are balanced. For example, heterologous Gα subunits may be expressed from a low copy (CEN ARS) vector containing the endogenous yeast GPA1 promoter and the GPA1 3' untranslated region. The minimum criterion, applied to a heterologous Gα subunit with respect to its ability to couple functionally to the yeast pheromone pathway, is that it complement a gpa1 genotype when expressed from the GPA1 promoter on low copy plasmids or from an integrated, single copy gene. In the work described in this application, all heterologous Gα subunits have been assayed in two biological systems. In the first assay heterologous Gα subunits are tested for an ability to functionally complement the growth arrest phenotype of *gpa1* strains. In the second assay the transcription of a *his3-HIS3* reporter gene is used to measure the extent to which the pheromone response pathway is activated, and hence the extent to which the heterologous Gα subunit sequesters the endogenous yeast Gβγ complex. Mammalian Gαs, Gαi2, Gαi3, Gαq, Gα11, Gα16, Gαo, Gαop, and Gαz from rat, murine or human origins were expressed from a low copy, CEN ARS vector containing the GPA1 promoter. Functional complementation of *gpa1* strains was not observed in either assay system with any of these full-length Gα constructs with the exception of rat and human Gαs.

*H. Chimeric Yeast Gβ subunits*

An alternative to the modification of a mammalian Gα subunit for improved signal transduction is the modification of the pertinent sites in the yeast Gβ or Gγ subunits. The principles discussed already with respect to Gα subunits apply, mutatis mutandis, to yeast Gβ or Gγ.

For example, it would not be unreasonable to target the yeast Step4 Gp subunit with cassette mutagenesis. Specifically, the region of Step4 that encodes several of the dominant negative, signaling-defective mutations would be an excellent target for cassette mutagenesis when looking for coupling of yeast Gp to specific mammalian G $\alpha$  subunits.

### *X. Peptide Libraries*

While others have engineered yeast cells to facilitate screening of exogenous drugs as receptor agonists and antagonists, the cells did not themselves produce both the drugs and the receptors. Yeast cells engineered to produce the receptor, but that do not produce the drugs themselves, are inefficient. To utilize them one must bring a sufficient concentration of each drug into contact with a number of cells in order to detect whether or not the drug has an action. Therefore, a microtiter plate well or test tube must be used for each drug. The drug must be synthesized in advance and be sufficiently pure to judge its action on the yeast cells. When the yeast cell produces the drug, the effective concentration is higher.

Peptide libraries are systems which simultaneously display, in a form which permits interaction with a target, a highly diverse and numerous collection of peptides. These peptides may be presented in solution (Houghten (1992) *Biotechniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (Ladner USP 5,223,409), spores (Ladner USP 4,099), plasmids (Cull et al. (1992) *Proc Natl Acad Sci USA* 89:1865-1869) or on phage (Scott and Smith (1990) *Science* 249:386-390); (Devlin (1990) *Science* 249:404-406); (Cwiria et al. (1990) *Proc. Natl. Acad. Sci.* 87:6378-6382); (Felic (1991) *J. Mol. Biol.* 222:301-310); (Ladner *supra*). Many of these systems are limited in terms of the maximum length of the peptide or the composition of the peptide (e.g., Cys excluded). Steric factors, such as the proximity of a support, may interfere with binding. Usually, the screening is for binding *in vitro* to an artificially presented target, not for activation or inhibition of a cellular signal transduction pathway in a living cell. While a cell surface receptor may be used as a target, the screening will not reveal whether the binding of the peptide caused an allosteric change in the conformation of the receptor.

The Ladner et al. patent, USSN 5,096,815, describes a method of identifying novel proteins or polypeptides with a desired DNA binding activity. Semi-random ("variegated") DNA encoding a large number of different potential binding proteins is introduced, in expressible form, into suitable host cells. The target DNA sequence is incorporated into a genetically engineered operon such that the binding of the protein or polypeptide will prevent expression of a gene product that is deleterious to the gene under selective conditions. Cells which survive the selective conditions are thus cells which express a protein which binds the target DNA. While it is taught that yeast cells may be used for testing, bacterial cells are preferred. The interactions between the protein and the target DNA occur only in the cell

(and then only in the nucleus), not in the periplasm or cytoplasm, and the target is a nucleic acid, and not a receptor protein. Substitution of random peptide sequences for functional domains in cellular proteins permits some determination of the specific sequence requirements for the accomplishment of function. Though the details of the recognition phenomena which operate in the localization of proteins within cells remain largely unknown, the constraints on sequence variation of mitochondrial targeting sequences and protein secretion signal sequences have been elucidated using random peptides (Lemire et al., *J. Biol. Chem.* (1989) 264, 20206 and Kaiser et al. (1987) *Science* 235:312, respectively).

The peptide library of the present invention takes the form of a cell culture, in which essentially each cell expresses one, and usually only one, peptide of the library. While the diversity of the library is maximized if each cell produces a peptide of a different sequence, it is usually prudent to construct the library so there is some redundancy. Depending on size, the combinatorial peptides of the library can be expressed as is, or can be incorporated into larger fusion proteins. The fusion protein can provide, for example, stability against degradation or denaturation, as well as a secretion signal if secreted. In an exemplary embodiment of a library for intracellular expression, e.g., for use in conjunction with intracellular target receptors, the polypeptide library is expressed as thioredoxin fusion proteins (see, for example, U.S. Patents 5,270,181 and 5,292,646; and PCT publication WO94/02502). The combinatorial peptide can be attached one the terminus of the thioredoxin protein, or, for short peptide libraries, inserted into the so-called active loop.

In one embodiment, the peptide library is derived to express a combinatorial library of polypeptides which are not based on any known sequence, nor derived from cDNA. That is, the sequences of the library are largely random. In preferred embodiments, the combinatorial polypeptides are in the range of 3-100 amino acids in length, more preferably at least 5-50, and even more preferably at least 10, 13, 15, 20 or 25 amino acid residues in length. Preferably, the polypeptides of the library are of uniform length. It will be understood that the length of the combinatorial peptide does not reflect any extraneous sequences which may be present in order to facilitate expression, e.g., such as signal sequences or invariant portions of a fusion protein.

In another embodiment, the peptide library is derived to express a combinatorial library of polypeptides which are based at least in part on a known polypeptide sequence or a portion thereof (not a cDNA library). That is, the sequences of the library is semi-random, being derived by combinatorial mutagenesis of a known sequence. See, for example, Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461. Accordingly, polypeptide(s) which are known ligands for a target receptor can be

mutagenized by standard techniques to derive a variegated library of polypeptide sequences which can further be screened for agonists and/or antagonists. For example, the surrogate ligand identified for FPR1-1, e.g., the Ser-Leu-Leu-Tip-Leu-Thr-Cys-Arg-Pro-Tip-Glu-Ala-Met peptide, can be mutagenized to generate a library of peptides with some relationship to the original tridecapeptide. This library can be expressed in a reagent cell of the present invention, and other receptor activators can be isolated from the library. This may permit the identification of even more potent FPR1-1 surrogate ligands.

Alternatively, the library can be expressed under conditions wherein the cells are in contact with the original tridecapeptide, e.g., the FPR1-1 receptor is being induced by that surrogate ligand. Peptides from an expressed library can be isolated based on their ability to potentiate the induction, or to inhibit the induction, caused by the surrogate ligand. The latter of course will identify potential antagonists of chemoattractant receptors. In still other embodiments, the surrogate ligand can be used to screen exogenous compound libraries (peptide and non-peptide) which, by modulating the activity of the identified surrogate, will presumably also similarly effect the native ligand's effect on the target receptor. In such embodiments, the surrogate ligand can be applied to the cells, though is preferably produced by the reagent cell, thereby providing an autocrine cell.

In still another embodiment, the combinatorial polypeptides are produced from a cDNA library. In a preferred embodiment of the present invention, the yeast cells collectively produce a "peptide library", preferably including at least  $10^3$  to  $10^7$  different peptides, so that diverse peptides may be simultaneously assayed for the ability to interact with the exogenous receptor. In an especially preferred embodiment, at least some peptides of the peptide library are secreted into the periplasm, where they may interact with the "extracellular" binding site(s) of an exogenous receptor. They thus mimic more closely the clinical interaction of drugs with cellular receptors. This embodiment optionally may be further improved (in assays not requiring pheromone secretion) by preventing pheromone secretion, and thereby avoiding competition between the peptide and the pheromone for signal peptidase and other components of the secretion system.

In the present invention, the peptides of the library are encoded by a mixture of DNA molecules of different sequence. Each peptide-encoding DNA molecule is ligated with a vector DNA molecule and the resulting recombinant DNA molecule is introduced into a host cell. Since it is a matter of chance which peptide encoding DNA molecule is introduced into a particular cell, it is not predictable which peptide that cell will produce. However, based on a knowledge of the manner in which the mixture was prepared, one may make certain statistical predictions about the mixture of peptides in the peptide library.

It is convenient to speak of the peptides of the library as being composed of constant



and variable residues. If the *n*th residue is the same for all peptides of the library, it is said to be constant. If the *n*th residue varies, depending on the peptide in question, the residue is a variable one. The peptides of the library will have at least one, and usually more than one, variable residue. A variable residue may vary among any of two to all twenty of the genetically encoded amino acids; the variable residues of the peptide may vary in the same or different manner. Moreover, the frequency of occurrence of the allowed amino acids at a particular residue position may be the same or different. The peptide may also have one or more constant residues.

There are two principal ways in which to prepare the required DNA mixture. In one method, the DNAs are synthesized a base at a time. When variation is desired, at a base position dictated by the Genetic Code, a suitable mixture of nucleotides is reacted with the nascent DNA, rather than the pure nucleotide reagent of conventional polynucleotide synthesis.

The second method provides more exact control over the amino acid variation. First, trinucleotide reagents are prepared, each trinucleotide being a codon of one (and only one) of the amino acids to be featured in the peptide library. When a particular variable residue is to be synthesized, a mixture is made of the appropriate trinucleotides and reacted with the nascent DNA. Once the necessary "degenerate" DNA is complete, it must be joined with the DNA sequences necessary to assure the expression of the peptide, as discussed in more detail below, and the complete DNA construct must be introduced into the yeast cell.

#### ***XI. Screening and Selection: Assays of Second Messenger Generation***

When screening for bioactivity of peptides, intracellular second messenger generation can be measured directly. A variety of intracellular effectors have been identified as being G-protein-regulated, including adenylyl cyclase, cyclic GMP, phosphodiesterases, phosphoinositidase C, and phospholipase A<sub>2</sub>. In addition, G proteins interact with a range of ion channels and are able to inhibit certain voltage-sensitive Ca<sup>++</sup> transients, as well as stimulating cardiac K<sup>+</sup> channels.

In one embodiment, the GTPase enzymatic activity by G proteins can be measured in plasma membrane preparations by determining the breakdown of  $\gamma^{32}\text{P}$  GTP using techniques that are known in the art (For example, see *Signal Transduction: A Practical Approach*, G. Milligan, Ed. Oxford University Press, Oxford England). When receptors that modulate cAMP are tested, it will be possible to use standard techniques for cAMP detection, such as competitive assays which quantitate [<sup>3</sup>H]cAMP in the presence of unlabelled cAMP.

Certain receptors stimulate the activity of phospholipase C which stimulates the breakdown of phosphatidylinositol 4,5, biphosphate to 1,4,5-IP<sub>3</sub> (which mobilizes

intracellular  $Ca^{++}$  and diacylglycerol (DAG) (which activates protein kinase C). Inositol lipids can be extracted and analyzed using standard lipid extraction techniques. DAG can also be measured using thin-layer chromatography. Water soluble derivatives of all three inositol lipids (IP1, IP2, IP3) can also be quantitated using radiolabelling techniques or HPLC.

The mobilization of intracellular calcium or the influx of calcium from outside the cell can be measured using standard techniques. The choice of the appropriate calcium indicator, fluorescent, bioluminescent, metallochromic, or  $Ca^{++}$ -sensitive microelectrodes depends on the cell type and the magnitude and time constant of the event under study (Borle (1990) *Environ Health Perspect* 84:45-56). As an exemplary method of  $Ca^{++}$  detection, cells could be loaded with the  $Ca^{++}$ -sensitive fluorescent dye fura-2 or indo-1, using standard methods, and any change in  $Ca^{++}$  measured using a fluorometer.

The other product of PIP2 breakdown, DAG can also be produced from phosphatidylcholine. The breakdown of this phospholipid in response to receptor-mediated signaling can also be measured using a variety of radiolabelling techniques.

The activation of phospholipase A2 can easily be quantitated using known techniques, including, for example, the generation of arachadonate in the cell.

In the case of certain receptors, it may be desirable to screen for changes in cellular phosphorylation. Such assay formats may be useful when the receptor of interest is a receptor tyrosine kinase. For example, yeast transformed with the FGF receptor and a ligand which binds the FGF receptor could be screened using colony immunoblotting (Lyons and Nelson (1984) *Proc. Natl. Acad. Sci. USA* 81:7426-7430) using anti-phosphotyrosine. In addition, tests for phosphorylation could be useful when a receptor which may not itself be a tyrosine kinase, activates protein kinases that function downstream in the signal transduction pathway. Likewise, it is noted that protein phosphorylation also plays a critical role in cascades that serve to amplify signals generated at the receptor. Multi-kinase cascades allow not only signal amplification but also signal divergence to multiple effectors that are often cell-type specific, allowing a growth factor to stimulate mitosis of one cell and differentiation of another.

One such cascade is the MAP kinase pathway that appears to mediate both mitogenic, differentiation and stress responses in different cell types. Stimulation of growth factor receptors results in Ras activation followed by the sequential activation of c-Raf, MEK, and p44 and p42 MAP kinases (ERK1 and ERK2). Activated MAP kinase then phosphorylates many key regulatory proteins, including p90RSK and Elk-1 that are phosphorylated when MAP kinase translocates to the nucleus. Homologous pathways exist in mammalian and yeast cells. For instance, an essential part of the *S. cerevisiae* pheromone signaling pathway is comprised of a protein kinase cascade composed of the products of the STE11, STE7, and

FUS3/KSS1 genes (the latter pair are distinct and functionally redundant). Accordingly, phosphorylation and/or activation of members of this kinase cascade can be detected and used to quantitate receptor engagement. Phosphotyrosine specific antibodies are available to measure increases in tyrosine phosphorylation and phospho-specific antibodies are commercially available (New England Biolabs, Beverly, MA).

Modified methods for detecting receptor-mediated signal transduction exist and one of skill in the art will recognize suitable methods that may be used to substitute for the example methods listed.

## *XII. Screening and Selection Using Reporter Gene Constructs*

In addition to measuring second messenger production, reporter gene constructs can be used. Reporter gene constructs are prepared by operatively linking a reporter gene with at least one transcriptional regulatory element. If only one transcriptional regulatory element is included it must be a regulatable promoter. At least one the selected transcriptional regulatory elements must be indirectly or directly regulated by the activity of the selected cell-surface receptor whereby activity of the receptor can be monitored via transcription of the reporter genes.

The construct may contain additional transcriptional regulatory elements, such as a FIRE sequence, or other sequence, that is not necessarily regulated by the cell surface protein, but is selected for its ability to reduce background level transcription or to amplify the transduced signal and to thereby increase the sensitivity and reliability of the assay.

Many reporter genes and transcriptional regulatory elements are known to those of skill in the art and others may be identified or synthesized by methods known to those of skill in the art. Reporter genes

A reporter gene includes any gene that expresses a detectable gene product, which may be RNA or protein. Preferred reporter genes are those that are readily detectable. The reporter gene may also be included in the construct in the form of a fusion gene with a gene that includes desired transcriptional regulatory sequences or exhibits other desirable properties.

Examples of reporter genes include, but are not limited to CAT (chloramphenicol acetyl transferase) (Alton and Vapnek (1979), Nature 282: 864-869) luciferase, and other enzyme detection systems, such as beta-galactosidase; firefly luciferase (deWet et al. (1987), Mol. Cell. Biol. 7:725-737); bacterial luciferase (Engelbrecht and Silverman (1984), PNAS 1: 4154-4158; Baldwin et al. (1984), Biochemistry 23: 3663-3667); alkaline phosphatase (Toh et al. (1989) Eur. J. Biochem. 182: 231-238, Hall et al. (1983) J. Mol. Appl. Gen. 2: 101), human placental secreted alkaline phosphatase (Cullen and Malim (1992) Methods in

Transcriptional control elements include, but are not limited to, promoters, enhancers, and repressor and activator binding sites. Suitable transcriptional regulatory elements may be derived from the transcriptional regulatory regions of genes whose expression is rapidly induced, generally within minutes, of contact between the cell surface protein and the effector protein that modulates the activity of the cell surface protein. Examples of such genes include, but are not limited to, the immediate early genes (see, Sheng et al. (1990) *Neuron* 4: 477-485), such as c-fos, Immediate early genes are genes that are rapidly induced upon binding of a ligand to a cell surface protein. The transcriptional control elements that are preferred for use in the gene constructs include transcriptional control elements from immediate early genes, elements derived from other genes that exhibit some or all of the characteristics of the immediate early genes, or synthetic elements that are constructed such that genes in operative linkage therewith exhibit such characteristics. The characteristics of preferred genes from which the transcriptional control elements are derived include, but are not limited to, low or undetectable expression in quiescent cells, rapid induction at the transcriptional level within minutes of extracellular stimulation, induction that is transient and independent of new protein synthesis, subsequent shut-off of transcription requires new protein synthesis, and mRNAs transcribed from these genes have a short half-life. It is not necessary for all of these properties to be present.

In the most preferred constructs, the transcriptional regulatory elements are derived from the c-fos gene.

The c-fos proto oncogene is the cellular homolog of the transforming gene of FBJ osteosarcoma virus. It encodes a nuclear protein that most likely involved in normal cellular growth and differentiation. Transcription of c-fos is transiently and rapidly activated by growth factors and by other inducers of other cell surface proteins, including hormones, differentiation-specific agents, stress, mitogens and other known inducers of cell surface proteins. Activation is protein synthesis independent. The c-fos regulatory elements include (see, Verma et al. (1987) *Cell* 51: a TATA box that is required for transcription initiation; two upstream elements for basal transcription, and an enhancer, which includes an element with dyad symmetry and which is required for induction by TPA, serum, EGF, and PMA.

The 20 bp transcriptional enhancer element located between - 317 and - 298 bp upstream from the c-fos mRNA cap site, which is essential for serum induction in serum starved NIH 3T3 cells. One of the two upstream elements is located at -63-57 and it resembles the consensus sequence for cAMP regulation.

Other promoters and transcriptional control elements, in addition to those described above, include the vasoactive intestinal peptide (VIP) gene promoter (cAMP responsive; Fink et al. (1988), *Proc. Natl. Acad. Sci.* 85:6662-6666); the somatostatin gene promoter (cAMP

responsive; Montminy et al. (1986), Proc. Natl. Acad. Sci. 83:6682-6686); the proenkephalin promoter (responsive to cAMP, nicotinic agonists, and phorbol esters; Comb et al. (1986), Nature 323:353-356); the phosphoenolpyruvate carboxy-kinase gene promoter (cAMP responsive; Short et al. (1986), J. Biol. Chem. 261:9721-9726); the NGFI-A gene promoter (responsive to NGF, cAMP, and serum; Changélian et al. (1989), Proc. Natl. Acad. Sci. 86:377-381); and others that may be known to or prepared by those of skill in the art.

In certain assays it may be desirable to use changes in growth in the screening procedure. For example, one of the consequences of activation of the pheromone signal pathway in wild-type yeast is growth arrest. If one is testing for an antagonist of a G protein-coupled receptor, this normal response of growth arrest can be used to select cells in which the pheromone response pathway is inhibited. That is, cells exposed to both a known agonist and a peptide of unknown activity will be growth arrested if the peptide is neutral or an agonist, but will grow normally if the peptide is an antagonist. Thus, the growth arrest response can be used to advantage to discover peptides that function as antagonists.

However, when searching for peptides which can function as agonists of G protein-coupled receptors, or other pheromone system proteins, the growth arrest consequent to activation of the pheromone response pathway is an undesirable effect since cells that bind peptide agonists stop growing while surrounding cells that fail to bind peptides will continue to grow. The cells of interest, then, will be overgrown or their detection obscured by the background cells, confounding identification of the cells of interest. To overcome this problem the present invention teaches engineering the cell such that: 1) growth arrest does not occur as a result of exogenous signal pathway activation (e.g., by inactivating the FAR1 gene); and/or 2) a selective growth advantage is conferred by activating the pathway (e.g., by transforming an auxotrophic mutant with a HIS3 gene under the control of a pheromone-responsive promoter, and applying selective conditions).

It is, of course, desirable that the exogenous receptor be exposed on a continuing basis to the peptides. Unfortunately, this is likely to result in desensitization of the pheromone pathway to the stimulus. For example, the mating signal transduction pathway is known to become desensitized by several mechanisms including pheromone degradation and modification of the function of the receptor, G proteins, and/or downstream elements of the pheromone signal transduction by the products of the SST2, STE50, AFR1 (Konopka, J.B. (1993) *Mol. Cell. Biol.* 13:6876-6888) and SGV1, MSG5, and SIG1 genes. Selected mutations in these genes can lead to hypersensitivity to pheromone and an inability to adapt to the presence of pheromone. For example, introduction of mutations that interfere with function into strains expressing heterologous G protein-coupled receptors constitutes a significant improvement on wild type strains and enables the development of extremely sensitive bioassays for compounds that interact with the receptors. Other mutations e.g.

STB50, sgvl, bar1, ste2, ste3, pik1, msg5, sig1, and aft1, have the similar effect of increasing the sensitivity of the bioassay. Thus desensitization may be avoided by mutating (which may include deleting) the SST2 gene so that it no longer produces a functional protein, or by mutating one of the other genes listed above.

If the endogenous homolog of the receptor is produced by the yeast cell, the assay will not be able to distinguish between peptides which interact with the endogenous receptor and those which interact with the exogenous receptor. It is therefore desirable that the endogenous gene be deleted or otherwise rendered nonfunctional.

In the case of receptors which modulate cyclic AMP, a transcriptional based readout can be constructed using the cyclic AMP response element binding protein, CREB, which is a transcription factor whose activity is regulated by phosphorylation at a particular serine (S133). When this serine residue is phosphorylated, CREB binds to a recognition sequence known as a CRE (cAMP Responsive Element) found to the 5' of promoters known to be responsive to elevated cAMP levels. Upon binding of phosphorylated CREB to a CRE, transcription from this promoter is increased.

Phosphorylation of CREB is seen in response to both increased cAMP levels and increased intracellular Ca levels. Increased cAMP levels result in activation of PKA, which in turn phosphorylates CREB and leads to binding to CRE and transcriptional activation. Increased intracellular calcium levels results in activation of calcium/calmodulin responsive kinase IV (CaM kinase IV). Phosphorylation of CREB by CaM kinase IV is effectively the same as phosphorylation of CREB by PKA, and results in transcriptional activation of CRE containing promoters.

Therefore, a transcriptional-based readout can be constructed in cells containing a reporter gene whose expression is driven by a basal promoter containing one or more CRE. Changes in the intracellular concentration of  $Ca^{++}$  (a result of alterations in the activity of the receptor upon engagement with a ligand) will result in changes in the level of expression of the reporter gene if: a) CREB is also co-expressed in the cell, and b) either the endogenous yeast CaM kinase will phosphorylate CREB in response to increases in calcium or if an exogenously expressed CaM kinase IV is present in the same cell. In other words, stimulation of PLC activity will result in phosphorylation of CREB and increased transcription from the CRE-construct, while inhibition of PLC activity will result in decreased transcription from the CRE-responsive construct.

As described in Bonni et al. (1993) *Science* 262:1575-1579, the observation that CNTF treatment of SK-N-MC cells leads to the enhanced interaction of STAT/p91 and STAT related proteins with specific DNA sequences suggested that these proteins might be key regulators of changes in gene expression that are triggered by CNTF. Consistent with this possibility is the finding that DNA sequence elements similar to the consensus DNA

sequence required for STAT/p91 binding are present upstream of a number of genes previously found to be induced by CNTF (e.g., Human c-fos, Mouse c-fos, Rat junB, Rat SOD-1, and CNTF). Those authors demonstrated the ability of STAT/p91 binding sites to confer CNTF responsiveness to a non-responsive reporter gene. Accordingly, a reporter construct for use in the present invention for detecting signal transduction through STAT proteins, such as from cytokine receptors, can be generated by using -71 to +109 of the mouse c-fos gene fused to the bacterial chloramphenicol acetyltransferase gene (-71fosCAT) or other detectable marker gene. Induction by a cytokine receptor induces the tyrosine phosphorylation of STAT and STAT-related proteins, with subsequent translocation and binding of these proteins to the STAT-RE. This then leads to activation of transcription of genes containing this DNA element within their promoters.

In preferred embodiments, the reporter gene is a gene whose expression causes a phenotypic change which is screenable or selectable. If the change is selectable, the phenotypic change creates a difference in the growth or survival rate between cells which express the reporter gene and those which do not. If the change is screenable, the phenotype change creates a difference in some detectable characteristic of the cells, by which the cells which express the marker may be distinguished from those which do not. Selection is preferable to screening in that it can provide a means for amplifying from the cell culture those cells which express a test polypeptide which is a receptor effector.

The marker gene is coupled to the receptor signaling pathway so that expression of the marker gene is dependent on activation of the receptor. This coupling may be achieved by operably linking the marker gene to a receptor-responsive promoter. The term "receptor-responsive promoter" indicates a promoter which is regulated by some product of the target receptor's signal transduction pathway.

Alternatively, the promoter may be one which is repressed by the receptor pathway, thereby preventing expression of a product which is deleterious to the cell. With a receptor repressed promoter, one screens for agonists by linking the promoter to a deleterious gene, and for antagonists, by linking it to a beneficial gene. Repression may be achieved by operably linking a receptor-induced promoter to a gene encoding mRNA which is antisense to at least a portion of the mRNA encoded by the marker gene (whether in the coding or flanking regions), so as to inhibit translation of that mRNA. Repression may also be obtained by linking a receptor-induced promoter to a gene encoding a DNA binding repressor protein, and incorporating a suitable operator site into the promoter or other suitable region of the marker gene.

In the case of yeast, suitable positively selectable (beneficial) genes include the following: *URA3*, *LYS2*, *HIS3*, *LEU2*, *TRP1*, *ADE1*, *2,3,4,5,7,8*; *ARG1*, *3*, *4*, *5*, *6*, *8*; *HIS1*, *4*, *5*, *ILV1*, *2*, *5*; *THR1*, *4*; *TRP2*, *3*, *4*, *5*; *LEU1*, *4*; *MET2*, *3,4,8,9,14,16,19*; *URA1*, *2,4,5,10*;

*HOM3,6; ASP3; CHOI; ARO 2,7; CYS3; OLE1; INO1,2,4; PR01,3* Countless other genes are potential selective markers. The above are involved in well-characterized biosynthetic pathways. The imidazoleglycerol phosphate dehydratase (IGP dehydratase) gene (*HIS3*) is preferred because it is both quite sensitive and can be selected over a broad range of expression levels. In the simplest case, the cell is auxotrophic for histidine (requires histidine for growth) in the absence of activation. Activation leads to synthesis of the enzyme and the cell becomes prototrophic for histidine (does not require histidine). Thus the selection is for growth in the absence of histidine. Since only a few molecules per cell of IGP dehydratase are required for histidine prototrophy, the assay is very sensitive.

In a more complex version of the assay, cells can be selected for resistance to aminotriazole (AT), a drug that inhibits the activity of IGP dehydratase. Cells with low, fixed level of expression of *HIS3* are sensitive to the drug, while cells with higher levels are resistant. The amount of AT can be selected to inhibit cells with a basal level of *HIS3* expression (whatever that level is) but allow growth of cells with an induced level of expression. In this case selection is for growth in the absence of histidine and in the presence of a suitable level of AT.

In appropriate assays, so-called counterselectable or negatively selectable genes may be used. Suitable genes include: *URA3* (orotidine-5'-phosphate decarboxylase; inhibits growth on 5-fluoroorotic acid), *LYS2* (2-aminoadipate reductase; inhibits growth on  $\alpha$ -amino adipate as sole nitrogen source), *CYH2* (encodes ribosomal protein L29; cycloheximide-sensitive allele is dominant to resistant allele), *CAN1* (encodes arginine permease; null allele confers resistance to the arginine analog canavanin), and other recessive drug-resistant markers.

In one example, the marker gene effects yeast cell growth. The natural response to signal transduction via the yeast pheromone system response pathway is for cells to undergo growth arrest. This is the preferred way to select for antagonists to a ligand/receptor pair that induces the pathway. An autocrine peptide antagonist would inhibit the activation of the pathway; hence, the cell would be able to grow. Thus, the *FAR1* gene may be considered an endogenous counterselectable marker. The *FAR1* gene is preferably inactivated when screening for agonist activity.

The marker gene may also be a screenable gene. The screened characteristic may be a change in cell morphology, metabolism or other screenable features. Suitable markers include beta-galactosidase (*Xgal*, *C12FDG*, Salmon-gal, Magenta-Gal (latter two from Biosynth Ag)), alkaline phosphatase, horseradish peroxidase, exo-glucanase (product of yeast *exb1* gene; nonessential, secreted), luciferase; bacterial green fluorescent protein; (human placental) secreted alkaline phosphatase (SEAP); and chloramphenicol transferase (CAT). Some of the above can be engineered so that they are secreted (although not  $\beta$ -galactosidase).



A preferred screenable marker gene is beta-galactosidase; yeast cells expressing the enzyme convert the colorless substrate Xgal into a blue pigment. Again, the promoter may be receptor-induced or receptor-inhibited.

### XIII. Genetic Markers in Yeast Strains

Yeast strains that are auxotrophic for histidine (HIS3) are known, see Struhl and Hill, (1987) *Mol. Cell. Biol.*, 7:104; Fasullo and Davis, *Mol. Cell. Biol.*, (1988) 8:4370. The HIS3 (imidazoleglycerol phosphate dehydratase) gene has been used as a selective marker in yeast. See Sikorski and Heiter, (1989) *Genetics*, 122:19; Struhl, et al., P.N.A.S. (1979) 76:1035; and, for FUS1-HIS3 fusions, see Stevenson, et al., (1992) *Genes Dev.*, 6:1293.

### LXX. Novel FPR1-Ligand

Yet another aspect of the invention pertains to a novel ligand for the orphan receptor, FPR1-1. As described in Example 8, a tridecapeptide having the sequence Ser-Leu-Leu-Trp-Leu-Thr-Cys-Arg-Pro-Trp-Glu-Ala-Met<sup>(SEQ ID NO:4)</sup> was identified from a polypeptide library on the basis of its ability to act as a surrogate ligand for FPR1-1.

Chemoattractants are important mediators of inflammation, they function to recruit phagocytic cells at the site of injury or infection. They also mediate granule secretion, superoxide generation and upregulation of cell surface adhesion molecules, for example MAC-1. Exemplary chemoattractants include the complement component C5a, interleukin 8, leukotriene B4 and platelet activating factor. Many of these substances participate in pathophysiological conditions such as anaphylaxis and septic shock. The identification of ligands for the orphan FPR1 receptor provides new opportunities for discovery of receptor agonists, that could potentially serve to enhance lymphocyte recruitment in immunocompromised patients, and for the discovery of receptor antagonists (described *supra*) that could prevent undesirable consequences of immune activation such as anaphylactic or septic shock.

The term "peptide" is used herein to refer to a chain of two or more amino acids or amino acid analogs (including non-naturally occurring amino acids), with adjacent amino acids joined by peptide (-NHCO-) bonds. Thus, the peptides of the present invention include oligopeptides, polypeptides, and proteins. Preferably, the peptides of the present invention include all or a portion of the S-L-L-W-L-T-C-R-P-W-E-A-M<sup>(SEQ ID NO:4)</sup> peptide, or a homolog thereof. The peptide (or peptidomimetic) is preferably at least 3 amino acid residues in length, though peptides of 4, 5, 7, 10, 13 or more residues in length are contemplated. For example, the sequence derived from the FPR1-1 surrogate ligand can be provided as part of a fusion protein. The minimum peptide length is chiefly dictated by the need to obtain sufficient

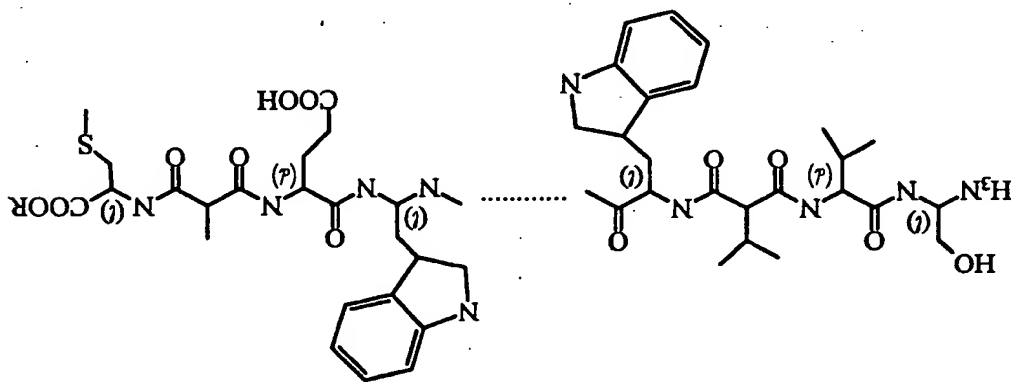
potency as an activator or inhibitor. Given the size of the peptide isolated in subject assay, smaller fragments of the tridecapeptide which retain receptor binding activity will be easily identified, e.g., by chemical synthesis of different fragments. The maximum peptide length will only be a function of synthetic convenience once an active peptide is identified.

The invention also provides for the generation of mimetics, e.g. peptide or non-peptide agents. Moreover, the present invention also contemplates variants of the subject polypeptide which may themselves be either agonistic or antagonistic of the S-L-L-W-L-T-C-R-P-W-E-A-M peptide. Thus, using such mutagenic techniques as known in the art, the determinants of S-L-L-W-L-T-C-R-P-W-E-A-M polypeptide which participate in FPR1-1 interactions can be elucidated. To illustrate, the critical residues of a subject polypeptide which are involved in molecular recognition of an FPR1-1 receptor can be determined and used to generate variant polypeptides which competitively inhibit binding of the authentic S-L-L-W-L-T-C-R-P-W-E-A-M peptide with that receptor. By employing, for example, scanning mutagenesis to map the amino acid residues of the polypeptide involved in binding the FPR1-1 receptor, peptide and peptidomimetic compounds can be generated which mimic those residues in binding to the receptor and which consequently can inhibit binding of an authentic ligand for the FPR1-1 receptor and interfere with the function of that receptor.

Moreover, as is apparent from the present and parent disclosures, mimetopes of the subject S-L-L-W-L-T-C-R-P-W-E-A-M peptide can be provided as non-hydrolyzable peptide analogs. For illustrative purposes, peptide analogs of the present invention can be generated using, for example, benzodiazepines (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gamma lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988, p.123), C-7 mimics (Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988, p.105), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985),  $\beta$ -turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231),  $\beta$ -aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun* 126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71), diamino ketones (Natarajan et al. (1984) *Biochem Biophys Res Commun* 124:141), and methyleneamino-modified (Roark et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988, p.134). Also, see generally, Session III: Analytic and synthetic methods, in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988).

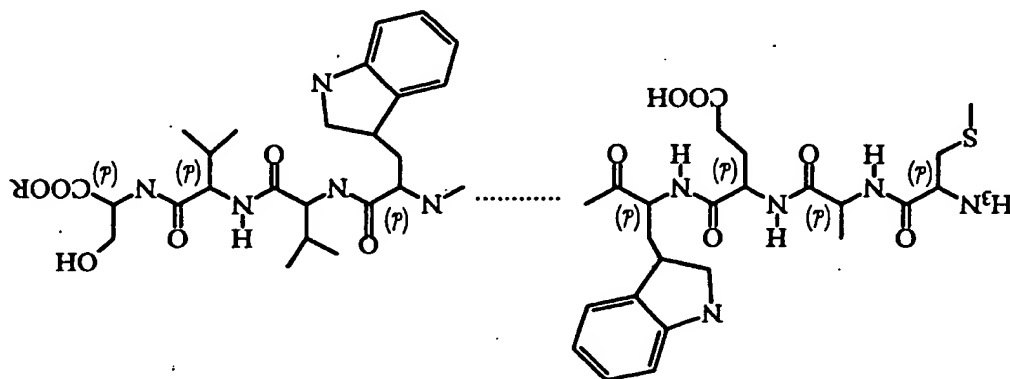
In an exemplary embodiment, the peptidomimetic can be derived as a retro-inverso

analog of the peptide. To illustrate, the S-L-L-W-L-T-C-R-P-W-E-A-M peptide can be generated as the retro-inverso analog:



Such retro-inverso analogs can be made according to the methods known in the art, such as that described by the Sisto et al. U.S. Patent 4,522,752. For example, the illustrated retro-inverso analog can be generated as follows. The geminal diamine corresponding to the serine analog is synthesized by treating a protected serine with ammonia under HOBt-DCC coupling conditions to yield the N-Boc amide, and then effecting a Hofmann-type rearrangement with I,1-bis-(trifluoroacetoxy)iodobenzene (TIB), as described in Radhakrishna et al. (1979) J. Org. Chem. 44:1746. The product amine salt is then coupled to a side-chain protected (e.g., as the benzyl ester) N-Fmoc D-Leu residue under standard conditions to yield the pseudodipeptide. The Fmoc (fluorenylmethoxycarbonyl) group is removed with piperidine in dimethylformamide, and the resulting amine is trimethylsilylated with bis(trimethylsilyl)acetamide (BSA) before condensation with suitably alkylated, side-chain protected derivative of Meldrum's acid, as described in U.S. Patent 5,061,811 to Pinnari et al., to yield the retro-inverso tripeptide analog S-L-L. The pseudotripeptide is then coupled with L-Trp under standard conditions to give the protected tetrapeptide analog. The protecting groups are removed to release the product, and the steps repeated to elongate the tetrapeptide to the full length peptide. It will be understood that a mixed peptide, e.g. including some normal peptide linkages, can be generated. As a general guide, sites which are most susceptible to proteolysis are typically altered, with less susceptible amide linkages being optional for mimetic switching. The final product, or intermediates thereof, can be purified by HPLC.

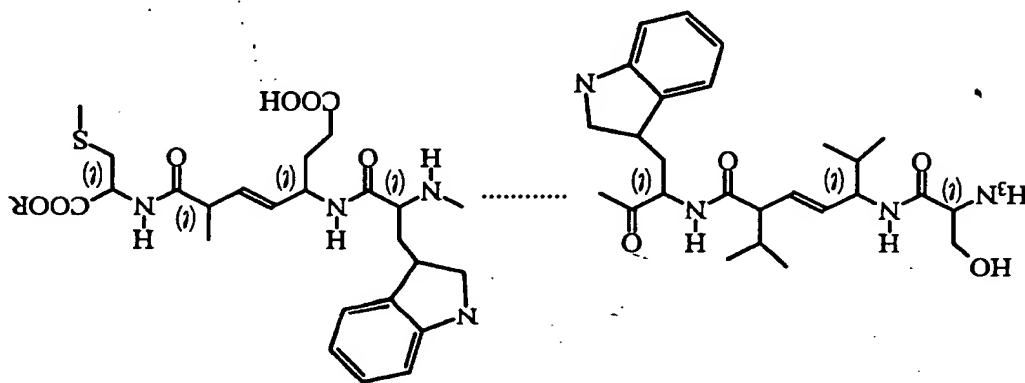
In another illustrative embodiment, the peptidomimetic can be derived as a retro-enatio analog of the peptide, such as the exemplary retro-enatio peptide derived for the illustrative S-L-L-W-L-T-C-R-P-W-E-A-M peptide:



NH<sub>3</sub>-(d) Met-(d) Ala-(d) Glu-(d) Trp.... (d) Trp-(d) Leu-(d) Ser

Retro-*enantiomer* analogs such as this can be synthesized using commercially available D-amino acids and standard solid- or solution-phase peptide-synthesis techniques. For example, in a preferred solid-phase synthesis method, a suitably amino-protected (t-butyloxycarbonyl, Boc) D-Serine residue (or analog thereof) is covalently bound to a solid support such as chloromethyl resin. The resin is washed with dichloromethane (DCM), and the Boc protecting group removed by treatment with TFA in DCM. The resin is washed and neutralized, and the next Boc-protected D-amino acid (D-Leu) is introduced by coupling with diisopropylcarbodiimide. The resin is again washed, and the cycle repeated for each of the remaining amino acids in turn (D-Leu, D-Trp etc). When synthesis of the protected retro-*enantiomer* peptide is complete, the protecting groups are removed and the peptide cleaved from the solid support by treatment with hydrofluoric acid/anisole/dimethyl sulfide/thioanisole. The final product is purified by HPLC to yield the pure retro-*enantiomer* analog.

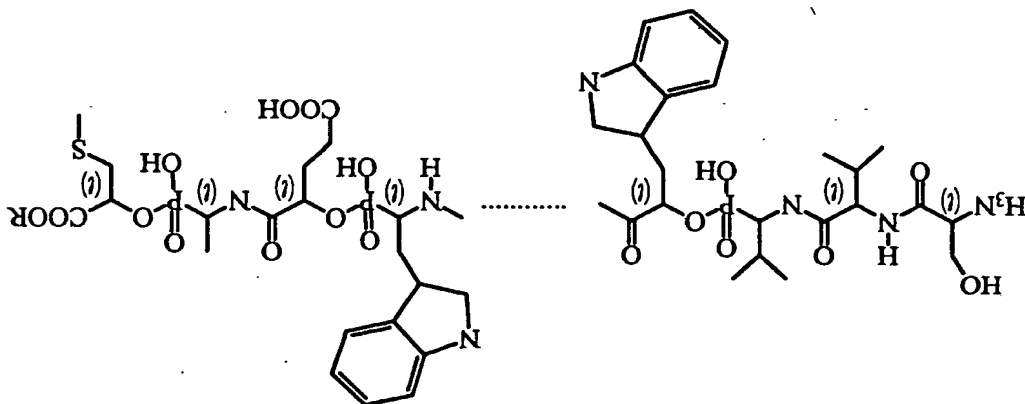
In still another illustrative embodiment, trans-olefin derivatives can be made for the subject polypeptide. For example, an exemplary olefin analog is derived for the illustrative S-L-L-W-L-T-C-R-P-W-E-A-M peptide:



The trans olefin analog of the subject peptide can be synthesized according to the method of

Y.K. Shue et al. (1987) Tetrahedron Letters 28:3225.

Still another class of peptidomimetic derivatives include the phosphonate derivatives, such as the partially phosphonate derived S-L-L-W-L-T-C-R-P-W-E-A-M peptide:



The synthesis of such phosphonate derivatives can be adapted from known synthesis schemes. See, for example, Loots et al. in Peptides: Chemistry and Biology, (Escom Science Publishers, Leiden, 1988, p. 118); Pettilo et al. in Peptides: Structure and Function (Proceedings of the 9th American Peptide Symposium, Pierce Chemical Co. Rockland, IL, 1985).

### Exemplification

The invention now being generally described will be more readily understood by reference to the following examples, which are included merely for purposes of illustration of certain aspects and embodiments of the present invention and are not intended to limit the invention.

### Example 1: Development of Autocrine Yeast Strains

In this example, we describe a pilot experiment in which haploid cells were engineered to be responsive to their own pheromones. (Note that in the examples, functional genes are capitalized and inactivated genes are in lower case.) For this purpose we constructed recombinant DNA molecules designed to:

- i. place the coding region of STE2 under the transcriptional control of elements which normally direct the transcription of STE3. This is done in a plasmid that allows the replacement of genomic STE3 of *S. cerevisiae* with sequences wherein the coding sequence

of *STE2* is driven by *STE3* transcriptional control elements.

ii. place the coding region of *STE3* under the transcriptional control of elements which normally direct the transcription of *STE2*. This is done in a plasmid which will allow the replacement of genomic *STE2* of *S. cerevisiae* with sequences wherein the coding sequence of *STE3* is driven by *STE2* transcriptional control elements.

The sequence of the *STE2* gene is known see Burkholder A.C. and Hartwell L.H. (1985), *Nuc. Acids Res.* 13, 8463; Nakayama N., Miyajima A., Arai K. (1985) *EMBO J.* 4, 2643.

A 4.3 kb BamHI fragment that contains the entire *STE2* gene was excised from plasmid YEp24-*STE2* (obtained from J. Thorne, Univ. of California) and cloned into pALTER (Protocols and Applications Guide, 1991, Promega Corporation, Madison, WI). An *SpeI* site was introduced 7 nucleotides (nts) upstream of the ATG of *STE2* with the following mutagenic oligonucleotide, using the *STE2* minus strand as template:

5'-GTTAAGAACCATATACCTAGATCAAAATGCTG 3'

A second *SpeI* site was simultaneously introduced just downstream of the *STE2* stop codon with the following mutagenic oligonucleotide:

5'-TGATCAAAATTACTAGTTGAAAAAGTAATTTCG 3'

The BamHI fragment of the resulting plasmid (Cadus 1096) containing *STE2* with *SpeI* sites immediately flanking the coding region, was then subcloned into the yeast integrating vector YIp19 to yield Cadus 1143.

The *STE3* sequence is also known (Nakayama N., Miyajima A., Arai K. (1985), *EMBO J.* 4, 2643; (Hagen D.C., McCaffrey G., Sprague G.F. (1986), *Proc. Natl. Acad. Sci.* 83, 1418. *STE3* was made available by Dr. J. Broach as a 3.1 kb fragment cloned into pBLUESCRIPT-KS II (Stratagene, 11011 North Torrey Pines Road, La Jolla, CA 92037). *STE3* was subcloned as a KpnI-XbaI fragment into both M13mp18 RF (to yield Cadus 1105 and pUC19 (to yield Cadus 1107). The two *SpeI* sites in Cadus 1107 were removed by digestion with *SpeI*, fill-in with DNA polymerase I Klenow fragment, and recircularization by blunt-end ligation. Single-stranded DNA containing the minus strand of *STE3* was obtained using Cadus 1105 and *SpeI* sites were introduced 9 nts upstream of the start codon and 3 nts downstream of the stop codon of *STE3* with the following mutagenic oligonucleotides, respectively:

5'-GGCAAAATAGTAAATTTTCATGTC 3'

5'-GGCCCTTAACACAGTGTGCGCATTAATTAC 3'

The mutagenesis was accomplished using the T7-GEN protocol of United States Biochemical (T7-GEN *In Vitro* Mutagenesis Kit, Descriptions and Protocols, 1991, United

States Biochemical, P.O. Box 22400, Cleveland, Ohio 44122). The replicative form of the resulting Cadus 1141 was digested with AflIII and KpnI, and the approximately 2 kb fragment containing the entire coding region of *STE3* flanked by the two newly introduced Spe I sites was isolated and ligated with the approximately 3.7 kb vector fragment of AflIII- and KpnI-digested Cadus 1107, to yield Cadus 1138. Cadus 1138 was then digested with XbaI and KpnI, and the *STE3*-containing 2.8 kb fragment was ligated into the XbaI- and KpnI-digested yeast integrating plasmid PRS406 (Sikorski, R.S. and Hieter, P. (1989), *Genetics* 122:19-27) to yield Cadus 1145.

The SpeI fragment of Cadus 1143 was replaced with the SpeI fragment of Cadus 1145 to yield Cadus 1147, in which the coding sequences of *STE3* are under the control of *STE2* expression elements. Similarly, the SpeI fragment of Cadus 1145 was replaced with the SpeI fragment of Cadus 1143 to yield Cadus 1148, in which the coding sequences of *STE2* are under the control of *STE3* expression elements. Using the method of pop-in/pop-out replacement (Rothstein, R. (1991) *Methods in Enzymology*, 194:281-301), Cadus 1147 was used to replace genomic *STE2* with the *ste2-STE3* hybrid in a MATa cell and Cadus 1148 was used to replace genomic *STE3* with the *ste3-STE2* hybrid in a MATa cell. Cadus 1147 and 1148 contain the selectable marker URA3.

Haploid yeast of mating type a which had been engineered to express *HIS3* under the control of the pheromone-inducible *FUS1* promoter were transformed with CADUS 1147, and transformants expressing URA3 were selected. These transformants, which express both *Ste2p* and *Ste3p*, were plated on 5-fluoroorotic acid to allow the selection of clones which had lost the endogenous *STE2*, leaving in its place the heterologous, integrated *STE3*. Such cells exhibited the ability to grow on media deficient in histidine, indicating autocrine stimulation of the pheromone response pathway.

Similarly, haploids of mating type  $\alpha$  that can express *HIS3* under the control of the pheromone-inducible *FUS1* promoter were transformed with CADUS 1148 and selected for replacement of their endogenous *STE3* with the integrated *STE2*. Such cells showed, by their ability to grow on histidine-deficient media, autocrine stimulation of the pheromone response pathway.

## Example 2: Strain Development

In this example, yeast strains are constructed which will facilitate selection of clones which exhibit autocrine activation of the pheromone response pathway. To construct appropriate yeast strains, we will use: the YIp-*STE3* and PRS-*STE2* knockout plasmids described above, plasmids available for the knockout of *FAR1*, *SST2*, and *HIS3*, and mutant strains that are commonly available in the research community. The following haploid strains will be constructed, using one-step or two-step knockout protocols described in *Mei*.

1.	MATa	ste3::STE2::ste3	far1	ssr12	FUS1::HIS3
2.	MATa	ste2::STE3::ste2	far1	ssr12	FUS1::HIS3
3.	MATa	ste3::STE2::ste3	far1	ssr12	mfa1 mfa2 FUS1::HIS3
4.	MATa	ste2::STE3::ste2	far1	ssr12	mfa1 mfa2 FUS1::HIS3
5.	MATa	bar1 far1-1 fus1-HIS3	ste14::TRP1	ura3 trp1 leu2 his3	
6.	MATa	mfa1 mfa2 far1-1	his3::fus1-HIS3	ste2-STE3	ura3 mel1 adel leu2

Strains 1 and 2 will be tested for their ability to grow on histidine-deficient media as a result of autocrine stimulation of their pheromone response pathways by the pheromones which they secrete. If these tests prove successful, strain 1 will be modified to inactivate endogenous *Mfa1* and *Mfa2*. The resulting strain 3, *MATa far1 ssr12 ste3::STE2::ste3 FUS1::HIS3 mfa1 mfa2*, should no longer display the selectable phenotype (i.e., the strain should be auxotrophic for histidine). Similarly, strain 2 will be modified to inactivate endogenous *Mfa1* and *Mfa2*. The resulting strain 4, *MATa far1 ssr12 ste2::STE3::ste2 FUS1::HIS3 mfa1 mfa2*, should be auxotrophic for histidine. The uses of strains 5 and 6 are outlined in Examples 3 and 4 below.

### Example 3: Peptide Library

In this example, a synthetic oligonucleotide encoding a peptide is expressed so that the peptide is secreted or transported into the periplasm.

i. The region of *Mfa1* which encodes mature  $\alpha$ -factor has been replaced via single-stranded mutagenesis with restriction sites that can accept oligonucleotides with AflIII and BglIII ends. Insertion of oligonucleotides with AflIII and BglIII ends will yield plasmids which encode proteins containing the *Mfa1* signal and leader sequences upstream of the sequence encoded by the oligonucleotides. The *Mfa1* signal and leader sequences should direct the processing of these precursor proteins through the pathway normally used for the transport of mature  $\alpha$ -factor.

The *Mfa1* gene, obtained as a 1.8 kb EcoRI fragment from pDA6300 (J. Thorne, Univ. of California) was cloned into PALTER in preparation for oligonucleotide-directed mutagenesis to remove the coding region of mature  $\alpha$ -factor while constructing sites for acceptance of oligonucleotides with AflIII and BglIII ends. The mutagenesis was accomplished using the minus strand as template and the following mutagenic oligonucleotide:

5'-CTAAGAAGAGGAGGATCTTTGCTTAAGCTCGAGATCTCGACTGATA-



ACAACAGTGTAG-3'

A HindIII site was simultaneously introduced 7 nts upstream of the Mf $\alpha$ 1 start codon with the oligonucleotide:

5' CATACACATATAAGCTTTAAAGAATGAG-3'

The resulting plasmid, Cadus 1214, contains a HindIII site 7 nts upstream of the Mf $\alpha$ 1 initiation codon, an AflIII site at the positions which encode the KEX2 processing site in the Mf $\alpha$ 1 leader peptide, and XhoI and BglIII sites in place of all sequences from the leader-encoding sequences up to and including the normal stop codon. The 1.5 kb HindIII fragment of Cadus 1214 therefore provides a cloning site for oligonucleotides to be expressed in yeast and secreted through the pathway normally travelled by endogenous  $\alpha$ -factor.

A sequence comprising the *ADC1* promoter and 5' flanking sequence was obtained as a 1.5 kb BamHI-HindIII fragment from pAAH5 (Armstrong, G. (1983) Academic Press, Inc., Meth. Enzymol. 101, 192-201 and ligated into the high copy yeast plasmid pRS426 (Christianson, T. W. et al. (1992) *Gene* 110:119-122) (see Figure 1). The unique XhoI site in the resulting plasmid was eliminated to yield Cadus 1186. The 1.5 Kb HindIII fragment of Cadus 1214 was inserted into HindIII-digested Cadus 1186; expression of sequences cloned into this cassette initiates from the *ADHI* promoter. The resulting plasmid, designated Cadus 1215, can be prepared to accept oligonucleotides with AflIII and BclI ends by digestion with those restriction endonucleases. The oligonucleotides will be expressed in the context of Mf $\alpha$ 1 signal and leader peptides (Figure 2).

Modified versions of Cadus 1215 were also constructed. To 30 improve the efficiency of ligation of oligonucleotides into the expression vector, Cadus 1215 was restricted with KpnI and religated to yield Cadus 1337. This resulted in removal of one of two HindIII sites. Cadus 1337 was linearized with HindIII, filled-in, and recircularized to generate Cadus 1338. To further tailor the vector for library construction, the following double-stranded oligonucleotide was cloned into AflIII- and BglIII-digested Cadus 1338:

5' TTAAGCGTGAGGAGCAAGCTTATCGATA oligo 062  
3' CGCACTCCGCTTCGAAATAGCTATCTAG oligo 063

The ClaI site is unique in the resulting vector, Cadus 1373. In Cadus 1373, the HindIII site that exists at the junction between the Mf $\alpha$  pro sequence and the mature peptide to be expressed by this vector was made unique. Therefore the HindIII site and the downstream BglIII site can be used to insert oligo-nucleotides encoding peptides of interest. These modifications of Cadus 1215 provide an alternative to the use of the AflIII site in the cloning of oligonucleotides into the expressions vector.

Two single-stranded oligonucleotide sequences (see below) are synthesized, annealed, and repetitively filled in, denatured, and reannealed to form double-stranded oligonucleotides that, when digested with AflIII and BclI, can be ligated into the polylinker of the expression vector, Cadus 1215. The two single-stranded oligonucleotides have the following sequences:

where N is any chosen nucleotide and n is any chosen integer. Yeast transformed with the resulting plasmids will secrete -- through the  $\alpha$ -factor secretory pathway -- peptides whose amino acid sequence is determined by the particular choice of N and n. Alternatively, the following single stranded oligonucleotides are used:

5'CTGGATGCGAAGACAGCTNNKNNKNNKNNKNNKNNKNNKNNKNNKNNKNN  
NNK TGATCAGTCTGTGACGC 3'

5' GCGTCACAGACTGATCA 3'

TGATCAGTCTGTGACGC

After fill-in using Tag DNA polymerase (Promega Corporation, Madison, Wisconsin), the double stranded product is restricted with BbsI and MboI and ligated to HindIII- and BglII-restricted Cadus 1373.

ii. The region of *MfaI* which encodes mature a-factor will be replaced via single stranded mutagenesis with restriction sites that can accept oligonucleotides with *XhoI* and *AflII* ends. Insertion of oligonucleotides with *XhoI* and *AflII* ends will yield plasmids which encode proteins containing the *MfaI* leader sequences upstream of the sequence encoded by the oligonucleotides. The *MfaI* leader sequences should direct the processing of these precursor proteins through the pathway normally used for the transport of mature a-factor.

*MfaI*, obtained as a BamHI fragment from pKK1 (provided by J. 30 Thorne and K. Kuchler), was ligated into the BamHI site of pALTER (Promega). Using the minus strand of *MfaI* as template, a HindIII site was inserted by oligonucleotide-directed mutagenesis just 5' to the *MfaI* start codon using the following oligonucleotide:

5' CCAATAAGTACAAAGCTTTCGAATAGAAATGCAACCATC

A second oligonucleotide was used simultaneously to introduce a short polylinker for later cloning of synthetic oligonucleotides in place of *MfaI* sequences. These *MfaI* sequences encode the C-terminal 5 amino acids of the 21 amino acid leader peptide through to the stop codon:

5'GCCGCTCCAAAGAAAGAACCTCGAGCTCGCTTAAGTTCTGCGTACAAACG-  
TTGTTTC 3'

The 1.6 kb HindIII fragment of the resulting plasmid, Cadus 1172, contains sequences encoding the *MfaI* start codon and the N-terminal 16 amino acids of the leader peptide, followed by a short polylinker containing *XhoI*, *SacI*, and *AflII* sites for insertion of oligonucleotides. The 1.6 kb HindIII fragment of Cadus 1172 was ligated into HindIII-digested Cadus 1186 (see above) to place expression of sequences cloned into this cassette under the control of the *ADHI* promoter. The *SacI* site in the polylinker was made unique by eliminating a second *SacI* site present in the vector. The resulting plasmid, designated Cadus 1239, can be prepared to accept oligonucleotides with *XhoI* and *AflII* ends by digestion with those restriction endonucleases for expression in the context of *MfaI* leader peptides (Figure 3).

Two single-stranded oligonucleotide sequences (see below) are synthesized, annealed, and repetitively filled in, denatured, and reannealed to form double-stranded oligonucleotides that, when digested with *AflII* and *BglII*, can be cloned into the polylinker of the expression vector, Cadus 1239. The two single-stranded oligonucleotides used for the cloning have the following sequences:

5' GG TAC TCG AGT GAA AAG AAG GAC AAC 3'

5' CG TAC TTA AGC AAT AAC ACA (NNN) <sup>n</sup> GTT GTC CTT CTT TTC ACT CGA

GTA CC 3'

where N is any chosen nucleotide and n is any chosen integer. Yeast transformed with the resulting plasmids will transport -- through the pathway normally used for the export of a-factor -- farnesylated, carboxymethylated peptides whose amino acid sequence is determined by the particular choice of N and n (Figure 3).

#### Example 4. Peptide Secretion/Transport.

This example demonstrates the ability to engineer yeast such that they secrete or transport oligonucleotide-encoded peptides (in this case their pheromones) through the pathways normally used for the secretion or transport of endogenous pheromones.

##### *Autocrine MATa strain CY588:*

A MATa strain designed for the expression of peptides in the context of MFa1 (i.e., using the MFa1 expression vector, Cadus 1215) has been constructed. The genotype of this strain, which we designate CY588, is *MATa bar1 far1 fus1-HIS3 ste14::TRP1 ura3 trp1 leu2 his3*. The *bar1* mutation eliminates the strain's ability to produce a protease that degrades  $\alpha$ -factor and that may degrade some peptides encoded by the cloned oligonucleotides; the *far1* mutation abrogates the arrest of growth which normally follows stimulation of the pheromone response pathway; an integrated FUS1-HIS3 hybrid gene provides a selectable signal of activation of the pheromone response pathway; and, finally, the *ste14* mutation lowers background of the FUS1-HIS3 readout. The enzymes responsible for processing of the MFa1 precursor in MATa cells are also expressed in MATa cells (Sprague and Thorne, in *The Molecular and Cellular Biology of the Yeast Saccharomyces: Gene Expression*, 1992, Cold Spring Harbor Press), therefore, CY588 cells should be able to secrete peptides encoded by oligonucleotides expressed from plasmid Cadus 1215.

A high transforming version (*tbl-1*) of CY588 was obtained by crossing CY1013 (CY588 containing an episomal copy of the STE14 gene) (*MATa bar1::hisGfar1-1 fus1-HIS3 ste14::TRP1 ura3 trp1 leu2 his3 [STE14 URA3 CEN4]*) to CY793 (*MATa~tbl-1 ura3 leu2 trp1 his3 fus1-HIS2 can1 ste114::TRP1 [FUS1 LEU2 2 $\mu$ ]*) and selecting from the resultant spores a strain possessing the same salient genotype described for CY588 (see above), and in addition the *tbl-1* allele, which confers the capacity for very high efficiency transformation by electroporation. The selected strain is CY1455 (*MATa bar1::hisGfar1-1 fus1-HIS3 ste14::TRP1 tbl-1 ura3 trp1 leu2 his3*).

##### *Secretion of peptides in the context of yeast $\alpha$ -factor:*

Experiments were performed to test: 1. the ability of Cadus 1215 to function as a vector for the expression of peptides encoded by synthetic oligonucleotides; 2. the suitability of the oligonucleotides, as designed, to direct the secretion of peptides through the  $\alpha$ -factor secretory pathway; 3. the capacity of CY588 to secrete those peptides; and 4. the ability of

CY588 to respond to those peptides that stimulate the pheromone response pathway by growing on selective media. These experiments were performed using an oligonucleotide which encodes the 13 amino acid  $\alpha$ -factor; i.e., the degenerate sequence (NNN) " in the oligonucleotide cloned into Cadus 1215 (see above) was specified (n=13) to encode this pheromone. CY588 was transformed with the resulting plasmid (Cadus 1219), and transformants selected on uracil-deficient medium were transferred to histidine-deficient medium supplemented with a range of concentrations of aminotriazole (an inhibitor of the *HIS3* gene product that serves to reduce background growth). The results demonstrate that the synthetic oligo-nucleotide, expressed in the context of MF $\alpha$ 1 by Cadus 1215, conferred upon CY588 an ability to grow on histidine-deficient media supplemented with aminotriazole. In summation, these data indicate that: 1. CY588 is competent for the secretion of a peptide encoded by the (NNN) " sequence of the synthetic oligonucleotide cloned into and expressed from Cadus 1215; and 2. CY588 can, in an autocrine fashion, respond to a secreted peptide which stimulates its pheromone response pathway, in this case by  $\alpha$ -factor binding to STE2.

Additional experiments were performed to test the utility of autocrine yeast strains in identifying agonists of the Ste2 receptor from among members of two semi-random  $\alpha$ -factor libraries,  $\alpha$ -Mid-5 and MF $\alpha$ -8.

#### *$\alpha$ -Mid-5 Library*

A library of semi-random peptides, termed the  $\alpha$ -Mid-5 library, was constructed. In this library, the N-terminal four amino acids and the C-terminal four amino acids of a 13 residue peptide are identical to those of native  $\alpha$ -factor while the central five residues (residues 5-9) are encoded by the degenerate sequence (NNQ) ". The following oligonucleotides were used in the construction of the  $\alpha$ -Mid-5 library:

(1) MF $\alpha$ Mbo, a 17 mer:

5' GCGTCACAGACTGATCA

(2) MID5ALF, a 71 mer:

5' GCCGTCAGTAAGCTTGGCATGGTGGNNQNNQNNQNNQNNQNNQMMQCCAGCCTATGTA  
CTGATC AGTCTGTGACGC

Sequenase (United States Biochemical Corporation, Cleveland, Ohio) was used to complete the duplex formed after annealing MF $\alpha$ Mbo to the MID5ALF oligonucleotide. In the MID5ALF sequence, N indicates a mixture of A, C, G, and T at ratios of 0.8:1:1.3:1; Q indicates a mixture of C and G at a ratio of 1:1.3. These ratios were employed to compensate for the different coupling efficiencies of the bases during oligonucleotide synthesis and were thus intended to normalize the appearance of all bases in the library. The double-stranded

oligonucleotide was restricted with HindIII and MboI and ligated to Cadus 1625 (see above); Cadus 1625 had been prepared to accept the semi-random oligonucleotides by restriction with HindIII and BglII.

The apparent complexity of the  $\alpha$ Mid-5 library is  $1 \times 10^7$ . This complexity is based on the number of bacterial transformants obtained with the library DNA versus transformants obtained with control vector DNA that lacks insert. Sequence analysis of six clones from the library demonstrated that each contained a unique insert.

To identify peptide members of the  $\alpha$ -mid-5 library that could act as agonists on the STB2 receptor, CY1455, a high transforming version of CY588, was electroporated to enhance uptake of  $\alpha$ -Mid-5 DNA. Transformants were selected on uracil-deficient (-Ura) synthetic complete medium and were transferred to histidine-deficient (-His) synthetic complete medium supplemented with 0.5mM or 1mM aminotriazole.

Yeast able to grow on -His + aminotriazole medium include (1) yeast which are dependent on the expression of an  $\alpha$ -factor variant agonist and (2) yeast which contain mutations that result in constitutive signalling along the pheromone pathway. Yeast expressing and secreting a variant STB2 receptor agonist have the ability to stimulate the growth on -His medium of surrounding CY 1455 cells that do not express such an agonist. Thus a recognizable formation (termed a "starburst") results, consisting of a central colony, growing by virtue of autocrine stimulation of the pheromone pathway, surrounded by satellite colonies, growing by virtue of paracrine stimulation of the pheromone pathway by the agonist peptide as that peptide diffuses radially from the central, secreting colony.

In order to identify the peptide sequence responsible for this "starburst" phenomenon, yeast were transferred from the center of the "starburst" and streaks were made on -Ura medium to obtain single colonies. Individual clones from -Ura were tested for the His+ phenotype on -His + aminotriazole plates containing a sparse lawn of CY1455 cells. Autocrine yeast expressing a peptide agonist exhibited the "starburst" phenotype as the secreted agonist stimulated the growth of surrounding cells that lacked the peptide but were capable of responding to it. Constitutive pheromone pathway mutants were capable of growth on -His + aminotriazole but were incapable of enabling the growth of surrounding lawn cells.

Alternatively, streaks of candidate autocrine yeast clones were made on plates containing 5-fluoroorotic acid (FOA) to obtain Ura segregants were retested on -His + aminotriazole for the loss of the His+ phenotype. Clones that lost the ability to grow on -His + aminotriazole after selection on FOA (and loss of the peptide-encoding plasmid) derived from candidate expressors of a peptide agonist. The plasmid was rescued from candidate clones and the peptide sequences determined. In addition, a plasmid encoding a putative Ste2 agonist was reintroduced into CY1455 to confirm that the presence of the plasmid

encoding the peptide agonist conferred the His<sup>+</sup> phenotype to CY1455.

By following the above protocol novel Stz2 agonists have been identified from the  $\alpha$ -Mid-5 library. Sequences of nine agonists follow, preceded by the sequence to the native  $\alpha$ -factor pheromone and by the oligonucleotide used to encode the native pheromone in these experiments. (Note the variant codons used in the  $\alpha$ -factor-encoding oligonucleotide for glutamine and proline in the C-terminal amino acids of  $\alpha$ -factor).

$\alpha$ -factor TGG CAT TGG TTG CAG CTA AAA CCT GGC CAA CCA ATG TAC  
encodes Trp His Trp Leu Gln Leu Lys Pro Gly Gln Pro Met Tyr

α-factor oligo:	encodes	Trp	His	TGG	TTG	CAG	CTA	AAA	CCT	GGC	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	M1	TGG	CAT	TGG	TTG	TCC	TTG	TGG	CCC	GGG	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	encodes	Trp	His	Trp	Leu	Ser	Leu	Ser	Pro	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	M2	TGG	CAT	TGG	TTG	TCC	CTG	GAC	GCT	GGC	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	encodes	Trp	His	Trp	Leu	Ser	Leu	Asp	Ala	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	M3	TGG	CAT	TGG	TTG	ACC	TTG	ATG	GCC	GGC	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	encodes	Trp	His	Trp	Leu	Ile	Leu	Met	Ala	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	M4	TGG	CAT	TGG	TTG	CAG	CTG	TGG	GCG	GGC	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	encodes	Trp	His	Trp	Leu	Gln	Leu	Ser	Ala	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC
	M5	TGG	CAT	TGG	TTG	AGG	TTG	CAG	TCC	GGC	Gln	CAG	CCT	Pro	Met	Tyr	TAC
encodes	Trp	His	Trp	Leu	Arg	Leu	Gln	Ser	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
M6	TGG	CAT	TGG	TTG	CGC	TTG	TCC	GCC	GGG	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
encodes	Trp	His	Trp	Leu	Arg	Leu	Gln	Ser	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
M7	TGG	CAT	TGG	TTG	TCC	CTC	GTC	CCG	GGG	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
encodes	Trp	His	Trp	Leu	Ser	Leu	Val	Pro	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
M8	TGG	CAT	TGG	TTG	TCC	CTG	TAC	CCC	GGG	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
encodes	Trp	His	Trp	Leu	Ser	Leu	Tyr	Pro	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
M9	TGG	CAT	TGG	TTG	CGG	CTG	CAG	CCC	GGG	Gln	CAG	CCT	Pro	Met	Tyr	TAC	
encodes	Trp	His	Trp	Leu	Arg	Leu	Gln	Pro	Gly	Gln	CAG	CCT	Pro	Met	Tyr	TAC	

The nine peptide agonists of the Stz2 receptor above were derived from one electroporation of CY1455 using 1  $\mu$ g of the  $\alpha$ -Mid-5 library DNA. Approximately  $3 \times 10^5$  transformants were obtained, representing approximately .03% of the sequences present in that library.

# *MFα-8 Library*

A semi-random α-factor library was obtained through synthesis of mutagenized α-factor oligonucleotides such that 1 in 10,000 peptide products were expected to be genuine α-factor. The mutagenesis was accomplished with doped synthesis of the oligonucleotides: each nucleotide was made approximately 68% accurate by synthesizing the following two oligos:

5' CTGATGCGAAGACTCAGCT (20 mer) (oligo060)

5' CGGATGATCA gta cat tgg gcc agg ttt tag ctg caa cca atg cca AGC TGA GTC TTC GCA TCC AG (69 mer) (oligo074)

The lower case letters indicate a mixture of 67% of that nucleotide and 11% of each of the other three nucleotides (e.g., t indicates 67% T and 11% A, 11% C, and 11% G). Note that digestion of the double-stranded oligo-nucleotide by FokI or BbsI will yield an identical 5' end that is compatible with HindIII ends.

Oligos 060 and 074 will form the following double-stranded molecule when annealed:

5'-CTGATGCGAAGACTCAGCT  
3'-GACCTACGCTTCTGAGTCGA acc gta acc aac gtc gat ttt gga ccg gtt ggt tac atg ACTAGTAGGC-5'

The duplex was repetitively filled-in using Taq DNA polymerase (Promega Corporation, Madison, Wisconsin). The double-stranded product was restricted with BbsI and BclI and ligated into HindIII- and BglII-digested Cadus 1373. The BglII/BclI joint creates a TGA stop codon for the termination of translation of the randomers. Using this approach, the MFα-5.8 library (a library of apparent low complexity based on PCR analysis of oligonucleotide insert frequency) was constructed.

To identify peptide members of the MFα-5.8 library that could act as agonists on the STB2 receptor, CY1455, a high transforming version of CY588, was electroporated to enhance uptake of MFα-5.8 DNA. Transformants were selected on uracil-deficient (-Ura) synthetic complete medium and were transferred to histidine-deficient (-His) synthetic complete medium supplemented with 1.0 mM or 2.5 mM aminotriazole. Yeast from colonies which were surrounded by satellite growth were transferred as streaks to -Ura medium to obtain single colonies. Yeast from single colonies were then tested for the His+ phenotype on -His + aminotriazole plates. Sequence analysis of seven of the plasmids



rescued from His<sup>+</sup> yeast revealed three unique  $\alpha$ -factor variants that acted as agonists on the STE2 receptor.

1.4 independent clones had the following sequence:

encoding      Trp His Trp Leu Gln Leu Thr Pro Gly Gln Pro Met Tyr  
TGG CAT TGG CTA CAG CTA ACG CCT GCG CAA CCA ATG TAC

2.2 independent clones had the following sequence:

encoding      Trp His Trp Leu Gln Leu Met Pro Gly Gln Pro Leu Tyr  
TGG CAT TGG CTG GAG CTT ATG CCT GCG CAA CCA TTA TAC

3.            encoding      Trp His Trp Met Gln Leu Arg Pro Gly Gln Pro Met Tyr  
TGG CAT TGG ATG GAG CTA AGA CCT GCG CAA CCA ATG TAC

#### *Autocrine Mata strain CY599:*

A MATA strain designed for the expression of peptides in the context of MFA1 (i.e., using the MFA1 expression vector, Cadus 1239) has been constructed. The genotype of this strain, designated CY599, is *MATA mfa1 mfa2 far-1 his3::fus1-HIS3 ste2-STE3 ura3 mel1 adel leu2*. In this strain, Cadus 1147 (see above) was used to replace STE2 with a hybrid gene in which the STE3 coding region is under the control of expression elements which normally drive the expression of STE2. As a result, the  $\alpha$ -factor receptor replaces the  $\alpha$ -factor receptor. The genes which encode  $\alpha$ -factor are deleted from this strain; the *far1* mutation abrogates the arrest of growth which normally follows stimulation of the pheromone response pathway; and the *FUS1-HIS3* hybrid gene (integrated at the *HIS3* locus) provides a selectable signal of activation of the pheromone response pathway. CY599 cells were expected to be capable of the transport of  $\alpha$ -factor or  $\alpha$ -factor-like peptides encoded by oligonucleotides expressed from Cadus 1239 by virtue of expression of the endogenous yeast transporter, Ste6.

#### *Transport of peptides by the yeast $\alpha$ -factor pathway:*

Experiments were performed to test: 1. the ability of Cadus 1239 to function as a vector for the expression of peptides encoded by synthetic oligonucleotides; 2. the suitability of the oligonucleotides, as designed, to direct the export of farnesylated, carboxymethylated peptides through the pathway normally used by  $\alpha$ -factor; 3. the capacity of CY599 to export these peptides; and 4. the ability of CY599 to respond to those peptides that stimulate the pheromone response pathway by growing on selective media. These tests were performed

using an oligonucleotide which encodes the 12 amino acid  $\alpha$ -factor; specifically, the degenerate sequence (NNN) " in the oligo-nucleotide cloned into Cadus 1239 (see above) (with  $n=12$ ) encodes the peptide component of  $\alpha$ -factor pheromone. CY599 was transformed with the resulting plasmid (Cadus 1220), and transformants selected on uracil-deficient medium were transferred to histidine-deficient medium supplemented with a range of concentrations of aminotriazole. The results demonstrate that the synthetic oligonucleotide, expressed in the context of MFA1 by Cadus 1220, conferred upon CY599 enhanced aminotriazole-resistant growth on histidine-deficient media. In summation, these data indicate: 1. Cadus 1220 and the designed oligonucleotide are competent to direct the expression and export of a farnesylated, carboxymethylated peptide encoded by the (NNN) " sequence of the synthetic oligonucleotide; and 2. CY599 can, in an autocrine fashion, respond to a farnesylated, carboxymethylated peptide that stimulates its pheromone response pathway, in this case signaling initiates as  $\alpha$ -factor binds to STE3.

#### Example 5. Proof of Concept

This example will demonstrate the utility of the autocrine system for the discovery of peptides which behave as functional pheromone analogues. By analogy, this system can be used to discover peptides that productively interact with any pheromone receptor surrogates.

CY588 (see strain 5, Example 2 above) will be transformed with CADUS 1215 containing oligonucleotides encoding random tridecapeptides for the isolation of functional  $\alpha$ -factor analogues. CY599 (see strain 6, Example 2 above) will be transformed with CADUS 1239 containing oligos of random sequence for the isolation of functional  $\alpha$ -factor analogues. Colonies of either strain which can grow on histidine-deficient media following transformation will be expanded for the preparation of plasmid DNA, and the oligo-nucleotide cloned into the expression plasmid will be sequenced to determine the amino acid sequence of the peptide which presumably activates the pheromone receptor. This plasmid will then be transfected into an isogenic strain to confirm its ability to encode a peptide which activates the pheromone receptor. Successful completion of these experiments will demonstrate the potential of the system for the discovery of peptides which can activate membrane receptors coupled to the pheromone response pathway.

Random oligonucleotides to be expressed by the expression plasmid CADUS 1215 will encode tridecapeptides constructed as

5' CGTGAAGCTTAAAGCGTGAGGAGGAGGCT(NNK)<sub>13</sub>TGATCATCCG, where N is any nucleotide, K is either T or G at a ratio of 40:60 (see Proc. Natl. Acad. Sci. 87:6378, 1990; *ibid* 89:5393, 1992), and the AIII and BcII sites are underlined. This oligonucleotide is designed such that: the AIII and BcII sites permit inserting the oligos into the AIII and BcII site of CADUS 1215, the HindIII site just 5' to the AIII site in the 5' end of the oligo allows

future flexibility with cloning of the oligos; the virtual repeat of GAGGCT and the GAGA repeats which are present in the wild-type sequence and which can form triple helixes are changed without altering the encoded amino acids. The random oligonucleotides described above will actually be constructed from the following two oligos:

5' CGTGAAGCTTAAGCGGTGAGGAGCAAGCT and  
5' CGGATGATCA(MNN)<sub>13</sub>AGCTTCTG,

where M is either A or C at a ratio of 40:60. The oligos will be annealed with one another and repetitively filled in, denatured, and reannealed (Kay et al, *Gene*, 1993). The double-stranded product will be cut with AflII and BclI and ligated into the AflII- and BglII-digested CADUS 1215. The BglII/BclI joint will create a TGA stop codon for termination of translation of the randommers. Because of the TA content of the AflI overhang, the oligos will be ligated to the AflII- and BglII-digested PADC-MFα at 4° C.

Random oligonucleotides to be expressed by the expression plasmid CADUS 1239 will encode monodecapetides constructed as

5' GGTACTCGAGTGAAGGAGGACAC(NNK)<sub>11</sub>TGTGTTATTGCTTAAGTACG,

where N is any nucleotide, K is either T or G at a ratio of 40:60 (see Proc. Natl. Acad. set 87:6378, 1990; *ibid* 89:5393, 1992). When cloned into the XhoI and AflII sites of CADUS 1239 the peptides expressed under the control of the ADHI promoter will contain the entire leader peptide of MFα, followed by 11 random amino acids, followed by triplets encoding CVIA (the C-terminal tetrapeptide of wild-type a-factor). Processing of the propeptide should result in the secretion of dodecapetides which contain 11 random amino acids followed by a C-terminal, farnesylated, carboxymethylated cysteine.

Using the procedure described above, the oligonucleotides for expression in CADUS 1239 will actually be constructed from the following two oligos:

5' GGTACTCGAGTGAAGGAGGACAC and  
5' CGTACTTAAGCAATAACACA(MNN)<sub>11</sub>GTTGTC,

where M is either A or C at a ratio of 40:60, and the XhoI and AflII sites are underlined.

# *Discovery of a-factor analogues from a random peptide library*

An optimized version of strain 6 (Example 2 above) was derived. This yeast strain, CY2012 (MATa *ste2-STE3 far1Δ1442 mfa2::LEU2 mfa2-lacZ fust-HIS3 ibi1-1 ura3 leu2 his3 trp1 suc2*), was constructed as follows. From a cross of CY570 (MATa *mfa2::LEU2 mfa2-lacZ ura3 trp1 his3Δ200 can1 leu2 fust-HIS3 [MFA1 URA3 2μ]* [Fus1Δ8-73 TRP1 CEN6]) by CY1624 (MATa *ibi1-1 fust-HIS3 trp1 ura3 leu2 his3 lys2-801 SUC+*), a spore was selected

(CY1877) of the following genotype: *MATa mfa1::LEU2 mfa2-lacZ fus1-HIS3 ibi1 ura3 leu2 his3 trp1 suc2*. This strain lacks both genes (*NFA1* and *MFA2*) encoding a-factor precursors, contains the appropriate pheromone pathway reporter gene (*fus1-HIS3*), and transforms by electroporation at high efficiency (*ibi1*). This strain was altered by deletion of the *FAR1* gene (with Cadus 1442; see Example 6), and replacement of *STE2* coding sequences with that of *STE3* (see Example 1) to yield CY2012.

This strain was transformed with plasmid DNA from a random a-factor library by electroporation and plated on 17 synthetic complete plates lacking uracil (-Ura), yielding approximately 10<sup>5</sup> Ura+ colonies per plate after 2 days at 30°C. These colonies were replica plated to histidine-deficient synthetic complete media (-His) containing 0.2 mM 3-aminotriazole and after three days at 30°C 35 His+ replicas were streaked to -Ura plates. The resultant colonies, 3 from each isolate, were retested for their His+ phenotype, and streaked to 5-fluoroorotic acid plates to obtain Ura segregants (lacking a library plasmid). Those Ura-segregants were tested for the loss of their His+ phenotype. Ten of the original isolates passed these tests; in two cases only one of the three Ura+ colonies purified from the isolate retained the His+ phenotype, but nevertheless subsequently segregated Ura His- colonies.

A single plasmid (corresponding to a bacterial colony) was obtained from each of the ten isolates, and reintroduced into CY2012. Eight of the ten plasmids passed the test of retaining the ability to confer the His+ phenotype on CY2012 (the two that failed correspond to the two isolates that were mentioned above, suggesting that these isolates contain at least one "irrelevant" plasmid). Sequencing of the randomized insert in the eight plasmids of interest revealed that four contain the sequence:

TAT GCT CTG TTT GTT CAT TTT TTT GAT ATT CCG  
Tyr Ala Leu Phe Val His Phe Phe Asp Ile Pro

two contain the sequence:

TTT AAG GGT CAG GTG CGT TTT GTG GTT GCT GCT

Phe Lys Gly Gln Val Arg Phe Val Val Leu Ala,

and two contain the sequence:

CTT ATG TCT CCG TCT TTT TTT TTG CCT GCG

Leu Met Ser Pro Ser Phe Phe Leu Pro Ala

Clearly, these sequences encode novel peptides, as the native a-factor sequence differs considerably:

Tyr Ile Ile Lys Gly Val Phe Trp Asp Pro Ala

The a-factor variants identified from random peptide libraries have utility as "improved" substrates of ABC transporters expressed in yeast. For example, identification of

a preferred substrate of human MDR, one that retains agonist activity on the pheromone receptor, would permit the establishment of robust yeast screens to be used in the discovery of compounds that affect transporter function.

#### Example 6: Functional Expression of a Mammalian G Protein-Coupled Receptor and Ligand in an Autocrine Yeast Strain.

This example details the following: (1) expression of human C5a receptor in yeast; (2) expression of the native ligand of this receptor, human C5a, in yeast; and (3) activation of the endogenous yeast pheromone pathway upon stimulation of the C5a receptor by C5a when both of these molecules are expressed within the same strain of autocrine yeast. Following the experimental data we outline the utility of autocrine strains of yeast that functionally express the human C5a receptor.

Human C5a is a 74 amino acid polypeptide that derives from the fifth component of complement during activation of the complement cascade; it is the most potent of the complement-derived anaphylatoxins. C5a is a powerful activator of neutrophils and macrophage functions including production of cytotoxic superoxide radicals and induction of chemotaxis and adhesiveness. In addition C5a stimulates smooth muscle contraction, induces degranulation of mast cells, induces serotonin release from platelets and increases vascular permeability. The C5a anaphylatoxin can also amplify the inflammatory response by stimulating the production of cytokines. As C5a is a highly potent inflammatory agent, it is a primary target for the development of antagonists to be used for intervention in a variety of inflammatory processes.

The C5a receptor is present on neutrophils, mast cells, macrophages and smooth muscle cells and couples through G proteins to transmit signals initiated through the binding of C5a.

#### Expression of the C5a Receptor

The plasmid pCDM8-C5aRc, bearing cDNA sequence encoding the human C5a receptor, was obtained from N. Gerard and C. Gerard (Harvard Medical School, Boston, MA) (Gerard and Gerard 1991). Sequence encoding C5a was derived from this plasmid by PCR using VENT polymerase (New England Biolabs Inc., Beverly MA), and the following primers:

#1-GGTGGAGGGTGCTCTAGAAGGAAGTTCACC  
#2-GCCGAGGAGACCAAGACCATGGAGCTCCCTCAATTATACCAACC

Primer #1 contains a single base-pair mismatch (underlined) to C5a receptor cDNA. It

introduces an XbaI site (in bold) 201 bp downstream from the TAG termination codon of the C5a receptor coding sequence. Primer #2 contains two mismatched bases and serves to create an NcoI site (in bold) surrounding the ATG initiator codon (double underlined). The second amino acid is changed from an aspartic acid to an asparagine residue. This is the only change in primary amino acid sequence from the wild type human C5a receptor.

The PCR product was restricted with NcoI and XbaI (sites in bold) and cloned into CADUS 1002 (YEP51Nco), a Gal10 promoter expression vector. The sequence of the entire insert was determined by dideoxy sequencing using multiple primers. The sequence between the NcoI and XbaI sites was found to be identical to the human C5a receptor sequence that was deposited in GenBank (accession #J053327) with the exception of those changes encoded by the PCR primers. The C5a receptor-encoding insert was transferred to CADUS 1289 (pLPX1), a PGK promoter expression vector, using the NcoI and XbaI sites, to generate the C5a receptor yeast expression clone, CADUS 1303.

A version of the C5a receptor which contains a yeast invertase signal sequence and a myc epitope tag at its amino terminus was expressed in Cadus 1270-transferred yeast under control of a GAL10 promoter. Plasmids encoding an untagged version of the C5a receptor and a myc-tagged derivative of FUS1 served as controls. The expression of the tagged receptor in yeast was confirmed by Western blot using the anti-myc monoclonal antibody 9E10. In the lane containing the extract from the Cadus 1270-transformant, the protein that is reactive with the anti-myc monoclonal antibody 9E10 was approximately 40 kD in size, as expected. Note that this receptor construct is not identical to the one used in the autocrine activation experiments. That receptor is not tagged, does not contain a signal sequence and is driven by the PGK promoter.

#### *Expression of the Ligand, C5a*

A synthetic construct of the sequence encoding C5a was obtained from C. Gerard (Harvard Medical School, Boston, MA). This synthetic gene had been designed as a FLAG-tagged molecule for the secretion from *E. coli* (Gerard and Gerard (1990) *Biochemistry* 29:9274-9281). The C5a coding region, still containing *E. coli* codon bias, was amplified using VENT polymerase (New England Biolabs Inc., Beverly MA) through 30 cycles using the following primers:

CSa5' = CCCCTTAAGCGGTGAGGCAAGGCTACTCTGCAAGAAGAGATC  
CSa3' = GAAGATCTTCAGCGGCGGAGTTGCGCATGTC

A PCR product of 257 bp was gel isolated, restricted with AflIII and BglII, and cloned into CADUS 1215 (an expression vector designed to express peptide sequences in the context of Mtx) to yield CADUS 1297. The regions of homology to the synthetic C5a gene are

underlined. The 5' primer also contains pre-pro  $\alpha$ -factor sequence. Upon translation and processing of the pre-pro  $\alpha$ -factor sequence, authentic human C5a should be secreted by yeast containing CADUS 1297. The insert sequence in CADUS 1297 was sequenced in both orientations by the dideoxy method and found to be identical to that predicted by the PCR primers and the published sequence of the synthetic C5a gene (Frank et al. (1988) *Methods in Enzymology* 162: 653-668).

Two sets of experiments, aside from the autocrine activation of yeast detailed below, demonstrated that CADUS 1297 can be used to express C5a in yeast.

1). C5a was immunologically detected in both culture supernatant and lysed cells using a commercially available enzyme-linked immunosorbent assay (ELISA)(Table 1). This assay indicated the concentration of C5a in the culture supernatant to be approximately 50 to 100 nM. In comparison, in data derived from mammalian cells, the binding constant of C5a to its receptor is 1 nM (Boulay et al.(1991) *Biochemistry* 30:2993-2999).

2). C5a expressed in yeast was shown to compete for binding with commercially obtained (Amersham Corporation, Arlington Heights, IL), radiolabeled C5a on induced HL60 cells.

#### *Activation of the Pheromone Response Pathway in Autocrine Yeast Expressing the Human C5a Receptor and Human C5a*

Activation of the yeast pheromone response pathway through the interaction of C5a with the C5a receptor was demonstrated using a growth read-out. The strain used for this analysis, CY455 (MAT $\alpha$  tbt1-1 ura3 leu2 trp1 his3 fus1-HIS3 can1 ste14::TRP1 ste3\*1156) contains the following significant modifications. A pheromone inducible HIS3 gene, fus1-HIS3, is integrated at the Fus1 locus. A hybrid gene containing sequence encoding the first 41 amino acids of GPA1 (the yeast G $\alpha$  subunit) fused to sequence encoding human G $\alpha$ 2a (minus codons for the N-terminal 33 amino acids) replaces GPA1 at its normal chromosomal location. The yeast STE14 gene is disrupted to lower the basal level of signaling through the pheromone response pathway. The yeast a-factor receptor gene, STE3, is deleted. The last two modifications are probably not essential, but appear to improve the signal-to-noise ratio. CY455 (MAT $\alpha$  tbt1-1 ura3 leu2 trp1 his3 fus1-HIS3 can1 ste14::TRP1 ste3\*1156) was transformed with the following plasmids:

Cadus 1289 + Cadus 1215 = Receptor Ligand<sup>-</sup> (R-L-)  
 Cadus 1303 + Cadus 1215 = Receptor<sup>+</sup> Ligand<sup>-</sup> = R+L-  
 Cadus 1289 + Cadus 1297 = Receptor<sup>+</sup> Ligand<sup>+</sup> = (R-L+)  
 Cadus 1303 + Cadus 1297 = Receptor<sup>+</sup> Ligand<sup>+</sup> = (R+L+)  
 Receptor refers to the human C5a receptor.  
 Ligand refers to human C5a.

Three colonies were picked from each transformation and grown overnight in media lacking leucine and uracil, at pH 6.8 with 25 mM PIPES (LEU URA pH6.8 with 25 mM PIPES). This media was made by adding 0.45 ml of sterile 1M KOH and 2.5 ml of sterile 1M PIPES pH 6.8 to 100 ml of standard SD LEU- URA- media. After overnight growth the pH of this media is usually acidified to approximately pH 5.5. Overnight cultures were washed once with 25 mM PIPES pH 6.8 and resuspended in an equal volume of media lacking leucine, uracil and histidine (LEU URA HIS pH 6.8 with 25 mM PIPES). The optical density at 600 nm of a 1/20 dilution of these cultures was determined and the cultures were diluted into 25 mM PIPES pH 6.8 to a final OD<sub>600</sub> of 0.2. A volume (5ul) of this dilution equivalent to 10,000 cells was spotted onto selective (HIS+ TRP- pH6.8) plates. Only those strains expressing both C5a and its receptor (R+L+) show growth on the selective plates which lack histidine. All test strains are capable of growth on plates containing histidine. The R+L+ strain will grow on plates containing up to 5 mM aminotriazole, the highest concentration tested.

For verification of pheromone pathway activation and quantification of the stimulation, the activity of the fus1 promoter was determined colorimetrically using a fus1-lacZ fusion in a similar set of strains. CY878 (MAT $\alpha$  tbt-1 fus1-HIS3 can1 ste4::trp::LYS2 ste3\*1156 gpa1(41)-Gal2) was used as the starting strain for these experiments. This strain is a trp1 derivative of CY455. The transformants for this experiment contained CADUS 1584 (PRS424-fus1-lacZ) in addition to the receptor and ligand plasmids. Four strains were grown overnight in SD LEU URA TRP pH6.8 with 50mM PIPES to an OD<sub>600</sub> of less than 0.8. Assay of  $\beta$ -galactosidase activity (Guarente 1983) in these strains yields the data shown in Figure 4. The coupling of the C5a receptor to G $\alpha$  chimeras is shown in Table 2.

#### Projected Uses of the Autocrine C5a Strains:

A primary use of the autocrine C5a strains will be in the discovery of C5a antagonists. Inhibitors of the biological function of C5a would be expected to protect against tissue damage resulting from inflammation in a wide variety of inflammatory disease processes including but not limited to: respiratory distress syndrome (Duchateau et al. (1984) *Am Rev Respir Dis* 130:1058); (Hammerschmidt et al. (1980) *Lancet* 1:947), septic lung injury (Olson et al. 1985) *Ann Surg* 202:771), arthritis (Banerjee et al. (1989) *J. Immunol* 142:2237), ischemic and post-ischemic myocardial injury (Weisman (1990) *Science* 146:249); (Crawford et al. (1988) *Circulation* 78:1449) and burn injury (Gelfand et al. (1982) *J. Clin Invest* 70:1170).

The autocrine C5a system as described can be used to isolate C5a antagonists as follows:



1. High throughput screens to identify antagonists of C5a. A straightforward approach involves screening compounds to identify those which inhibit growth of the R+L+ strain described above in selective media but which do not inhibit the growth of the same strain or of a R+L- strain in non-selective media. The counterscreen is necessary to eliminate from consideration those compounds which are generally toxic to yeast. Initial experiments of this type have led to the identification of compounds with potential therapeutic utility.

## 2. Identification of antagonists using negative selection.

Replacement of the fus1-HIS3 read-out with one of several negative selection schemes (fus1-URA3/FOA, fus1-GAL1/galactose or deoxygalactose, Far1 ss12 or other mutations that render yeast supersensitive for growth arrest) would generate a test system in which the presence of an antagonist would result in the growth of the assay strain. Such an approach would be applicable to high-throughput screening of compounds as well as to the selection of antagonists from random peptide libraries expressed in autoclone yeast. Optimization of screens of this type would involve screening the R+L+ strain at a concentration of aminotriazole which ablates growth of the R+L- strain (we are currently using 0.6 to 0.8 mM) and counterscreening the R+L- strain at a concentration of aminotriazole which gives an identical growth rate (we are using 0.14 mM). In addition, the system could employ one of several colorimetric, fluorescent or chemiluminescent readouts. Some of the genes which can be fused to the fus1 promoter for these alternate read-outs include lacZ (colorimetric and fluorescent substrates), glucuronidase 20 (colorimetric and fluorescent substrates), phosphatases (e.g. PHO3, PHO5, alkaline phosphatase; colorimetric and chemiluminescent substrates), green protein (endogenous fluorescence), horse radish peroxidase (colorimetric), luciferase (chemiluminescence).

The autoclone C5a strains have further utility as follows:

3. In the identification of novel C5a agonists from random peptide libraries expressed in autoclone yeast. Novel peptide agonists would contribute to structure/function analyses used to guide the rational design of C5a antagonists.

## 4. In the identification of receptor mutants.

Constitutively active, that is, ligand independent, receptors may be selected from highly mutagenized populations by growth on selective media. These constitutively active receptors may have utility in permitting the mapping of the sites of interaction between the receptor and the G-protein. Identification of those sites may be important to the rational

design of drugs to block that interaction. In addition, receptors could be selected for an ability to be stimulated by some agonists but not others or to be resistant to antagonist. These variant receptors would aid in mapping sites of interaction between receptor and agonist or antagonist and would therefore contribute to rational drug design efforts.

5. In the identification of molecules that interact with Gα12.

Compounds or peptides which directly inhibit GDP exchange from Gα12 would have the same effect as C5a antagonists in these assays. Additional information would distinguish inhibitors of GDP exchange from C5a antagonists. This information could be obtained through assays that determine the following:

1. inhibition by test compounds of Gα12 activation from other receptors,
2. failure of test compounds to compete with radiolabeled C5a for binding to the C5a receptor,
3. failure of test compounds to inhibit the activation of other Gα subunits by C5a, and
4. inhibition by test compounds of signalling from constitutively active versions of C5a, or other, receptors.

Example 7: Construction of Hybrid Gα Genes Constructed of two sets of chimeric

yeast/mammalian Gα genes, GPA<sub>41</sub>-Gα and GPA<sub>Bam</sub>-Gα.

The Gα subunit of heterotrimeric G proteins must interact with both the βγ complex and the receptor. Since the domains of Gα required for each of these interactions have not been completely defined and since our final goal requires Gα proteins that communicate with a mammalian receptor on one hand and the yeast βγ subunits on the other, we desired to derive human-yeast chimeric Gα proteins with an optimized ability to perform both functions. From the studies reported here we determined that inclusion of only a small portion of the amino terminus of yeast Gα is required to couple a mammalian Gα protein to the yeast βγ subunits. It was anticipated that a further benefit to using these limited chimeras was the preservation of the entire mammalian domain of the Gα protein believed to be involved in receptor contact and interaction. Thus the likelihood that these chimeras would retain their ability to interact functionally with a mammalian receptor expressed in the same yeast cell was expected to be quite high.

Plasmid constructions.

PRS416-GPA1 (Cadus 1069). An XbaI - SacI fragment encoding the entire GPA1 promoter region, coding region and approximately 250 nucleotides of 3' untranslated region was excised from 10 YCplac111-GPA1 (from S. Reed, Scripps Institute) and cloned into

YEp vector pRS416 (Sikorski and Hieter, Genetics 122: 19 (1989)) cut with XbaI and SacI.

**Site-directed mutagenesis of *GPAI* (Cadus 1075, 1121 and 1122).** A 1.9 kb EcoRI fragment containing the entire *GPAI* coding region and 200 nucleotides from the 5' untranslated region was cloned into EcoRI cut, phosphatase-treated pALTER-1 (Promega) and transformed by electroporation (Biorad Gene Pulser) into DH5α<sup>r</sup> bacteria to yield Cadus 1075. Recombinant phagemids were rescued with M13KO7 helper phage and single stranded recombinant DNA was extracted and purified according to the manufacturer's specifications. A new NcoI site was introduced at the initiator methionine of *GPAI* by oligonucleotide directed mutagenesis using the synthetic oligonucleotide:

5' GATATATTAAAGGTAAGGAACCATGGGGTGTACAGTAG 3'

Positive clones were selected in ampicillin and several independent clones were sequenced in both directions across the new NcoI site at +1. Two clones containing the correct sequences were retained as Cadus 1121 and 1122.

**Construction of a *GPAI*-based expression vector (Cadus 1127).** The vector used for expression of full length and hybrid mammalian Gα proteins in yeast, Cadus 1127, was constructed in the following manner. A 350 nucleotide fragment spanning the 3' untranslated region of *GPAI* was amplified with Taq polymerase (AmpliTaq; Perkin Elmer) using the oligonucleotide primers A (5' CGAGGCTCGAGGGAACGATATAAAGTAGTG 3') and B (5' GCGCGGTACCAAGCTTCAATTCGAGATATACCC 3'). The 350 nucleotide product was purified by gel electrophoresis using GeneClean II (Bio101) and was cloned directly into the pCRII vector by single nucleotide overlap TA cloning (InVitrogen). Recombinant clones were characterized by restriction enzyme mapping and by dideoxynucleotide sequencing. Recombinant clones contained a novel XhoI site 5' to the authentic *GPAI* sequence and a novel KpnI site 3' to the authentic *GPAI* sequence donated respectively by primer A and primer B.

The NotI and SacI sites in the polylinker of Cadus 1013 (pRS414) were removed by restriction with these enzymes followed by filling in with the Klenow fragment of DNA polymerase I and blunt end ligation to yield Cadus 1092. The 1.4 kb PstI - EcoRI 5' fragment of *GPAI* from YCplac111-*GPAI* containing the *GPAI* promoter and 5' untranslated region of *GPAI* was purified by gel electrophoresis using GeneClean (Bio101) and cloned into PstI - EcoRI restricted Cadus 1013 to yield Cadus 1087. The PCR amplified XhoI - KpnI fragment encoding the 3' untranslated region of *GPAI* was excised from Cadus 1089 and cloned into XhoI - KpnI restricted Cadus 1087 to yield Cadus 1092. The NotI

and SacI sites in the polylinker of Cadus 1092 were removed by restriction with these enzymes, filling in with the Klenow fragment of DNA polymerase I, and blunt end ligation to yield Cadus 1110. The region of Cadus 1122 encoding the region of *GPAI* from the *EcoRI* site at -200 to +120 was amplified with Vent DNA polymerase (New England Biolabs, Beverly, MA) with the primers

5' CCCGAATCCACCAATTCTTACG 3' and  
5' GCGGCGTCCGACGCCGCCGTAACAGT 3'.

The amplified product, bearing an *EcoRI* site at its 5' end and novel *SacI*, *NciI* and *Sall* sites at its 3' end was restricted with *EcoRI* and *Sall*, gel purified using GeneClean II (Bio101), and cloned into *EcoRI* and *Sall* restricted Cadus 1110 to yield Cadus 1127. The DNA sequence of the vector between the *EcoRI* site at -200 and the *KpnI* site at the 3' end of the 3' untranslated region was verified by restriction enzyme mapping and dideoxynucleotide DNA sequence analysis.

**PCR amplification of *GPA<sub>41</sub>*-Gα proteins and cloning into Cadus 1127.** cDNA clones encoding the human Gα subunits *Gαs*, *Gαi2*, *Gαi3*, and *S. cerevisiae GPAI* were amplified with Vent thermostable polymerase (New England Biolabs, Beverly, MA). The primer pairs used in the amplification are as follows:

*Gαs* Primer 1 : 5'CTGCTGGAGCTCCGCCCTGCTGCTGGTGGTGGAG3' (*SacI* 5')  
Primer 2 : 5'CTGCTGGTCCGACGGCGCGGGGGGTTCTTCTTAGAAGCAGC3' (*Sall* 3')  
Primer 3 : 5'GGGCTCGAGCCCTTCTTAGAGCAGCTCGTAC3' (*XhoI* 3')

*Gαi2* Primer 1 : 5'CTGCTGGAGCTCAAGTTGCTGCTGCTGTTGGTGGTGGG3' (*SacI* 5')  
Primer 2 : 5'CTGCTGGTCCGACGGCGCGGGCGCGGGTCCAGAAGAGGCCGCGGTCC3' (*Sall* 3')

Primer 3 : 5'GGGCTCGAGCCCTCAGAAGAGGCCGCCAGTC3' (*XhoI* 3')  
*Gαi3* Primer 1: 5'CTGCTGGAGCTCAAGCTGCTGCTGCTGCTGCTGAG3' (*SacI* 5')  
Primer 2 : 5'CTGCTGGTCCGACGGCGCGGGCGCGGGTCCATCAGTCCATCCTCTCAATAAAGTC3' (*Sall* 3')  
Primer 3 : 5'GGGCTCGAGCATGCTTCTCAATAAAGTCCAC3' (*XhoI* 3')

After amplification, products were purified by gel electrophoresis using GeneClean II (Bio101) and were cleaved with the appropriate restriction enzymes for cloning into Cadus 1127.

The hybrid *GPA<sub>41</sub>*-Gα subunits were cloned via a *SacI* site introduced at the desired position near the 5' end of the amplified genes and a *Sall* or *XhoI* site introduced in the 3'

untanslated region. Ligation mixtures were electroporated into competent bacteria and plasmid DNA was prepared from 50 cultures of ampicillin resistant bacteria.

**Construction of Integrating Vectors Encoding GPA<sub>41</sub>-G<sub>α</sub> Subunits.** The coding region of each GPA<sub>41</sub>-G<sub>α</sub> hybrid was cloned into an integrating vector (pRS406 = URA3 Amp<sup>r</sup>) using the BssHII sites flanking the polylinker cloning sites in this plasmid. Cadus 1011 (pRS406) was restricted with BssHII, treated with shrimp alkaline phosphatase as per the manufacturer's specifications, and the linearized vector was purified by gel electrophoresis. Inserts from each of the GPA<sub>41</sub>-G<sub>α</sub> hybrids were excised with BssHII from the parental plasmid, and subcloned into gel purified Cadus 1011.

**Construction of GPA<sup>BAM</sup>-G<sub>α</sub> Constructs.** A novel BamHI site was introduced in frame into the GPA1 coding region by PCR amplification using Cadus 1179 (encoding a wildtype GPA1 allele with a novel NcoI site at the initiator methionine) as the template, VENT polymerase, and the following primers: Primer A = 5' GCATCCATCAATATCCAG 3' and Primer B = 5' GAAACAATGGA - TCACCTTCTAC 3'. The 1.1 kb PCR product was gel purified with GeneClean II (Bio101), restricted with NcoI and BamHI and cloned into NcoI-BamHI cut and phosphatased Cadus 1605. The sequence of Cadus 1605 was verified by restriction analysis and dideoxy-sequencing of double-stranded templates. Recombinant GPA<sup>BAM</sup>-G<sub>α</sub> hybrids of G<sub>α</sub>s, G<sub>α</sub>12, and G<sub>α</sub>16 were generated. Construction of Cadus 1855 encoding recombinant GPA<sup>BAM</sup>-G<sub>α</sub> 16 serves as a master example: construction of the other hybrids followed an analogous cloning strategy. The parental plasmid Cadus 1617, encoding native G<sub>α</sub>16, was restricted with NcoI and BamHI, treated with shrimp alkaline phosphatase as per the manufacturer's specifications and the linearized vector was purified by gel electrophoresis. Cadus 1605 was restricted with NcoI and BamHI and the 1.1 kb fragment encoding the amino terminal 60% of GPA1 with a novel BamHI site at the 3' end was cloned into the NcoI- and BamHI-restricted Cadus 1617. The resulting plasmid encoding the GPA<sup>BAM</sup>-G<sub>α</sub> 16 hybrid was verified by restriction analysis and assayed in tester strains for an ability to couple to yeast G<sub>β</sub>γ and thereby suppress the *gpa1* null phenotype. Two additional GPA<sup>BAM</sup>-G<sub>α</sub> hybrids, GPA<sup>BAM</sup>-G<sub>α</sub>s and GPA<sup>BAM</sup>-G<sub>α</sub>12, described in this application were prepared in an analogous manner using Cadus1606 as the parental plasmid for the construction of the GPA<sup>BAM</sup>-G<sub>α</sub> 12 hybrid and Cadus 1181 as the parental plasmid for the construction of the GPA<sup>BAM</sup>-G<sub>α</sub>s hybrid.

**Coupling by chimeric G<sub>α</sub> proteins.** The G<sub>α</sub> chimeras described above were tested for the ability to couple a mammalian G protein-coupled receptor to the pheromone response pathway in yeast. The results of these experiments are outlined in Table 3. Results obtained using GPA1<sub>41</sub>-G<sub>α</sub>12 to couple the human C5a receptor to the pheromone response pathway in

autocrine strains of yeast are disclosed in above.

#### Example 8: Screening for Modulators of G-alpha activity

Screens for modulators of Gα activity may also be performed as shown in the following examples for illustration purposes, which are intended to be non-limiting.

Strains CY4874 and CY4877 are isogenic but for the presence of Q205L mutation in the cloned  $G\alpha_2$  gene cloned into plasmid 1. Strains CY4901 and CY4904 each have a chromosomally integrated chimeric Gα fusion comprising 41 amino acids of gpal at the N terminus of the human  $G\alpha_{12}$  gene and are isogenic but for the presence of a constitutively activating mutation in the C5a receptor gene of CY4901. Strain CY5058 is a gpal mutant which carries only the yeast Gβγ subunits and no Gα subunit. This strain is a control strain to demonstrate specificity of action on the Gα subunit.

# I. Suppression of Activation by Mutation of $G\alpha$

The Q205L mutation is a constitutively activated GTPase deficient mutant of the human  $G\alpha_{12}$  gene. Antagonist compounds, chemicals or other substances which act on  $G\alpha_{12}$  can be recognized by their action to reduce the level of activation and thus reduce the signal from the *fus1-lacZ* reporter gene on the second plasmid (Plasmid 2).

## A. GTPase $G\alpha_{12}$ Mutants

test component =  $gpa_{41}-G\alpha_{12}$  (Q205L)  
control component =  $gpa_{41}-G\alpha_{12}$

As well as the CY4874 and CY4877 constructs detailed above, similar strains with *fus1-His3* or *fus2-CAN-1* growth readouts may also be used. The *fus1-His3* strains are preferred for screening for agonists and the *fus2-CAN-1* strains are preferred for antagonist screens.

Readout	test	effect of $G\alpha_{12}$ -antagonist	control
<i>fus1-His3</i>	strain CY4868	inhibit growth of -His +AT (Aminotriazole)	strain CY4871
<i>fus1-lacZ</i>	CY4874	reduce $\beta$ -gal activity	CY4877
<i>fus2-CAN1</i>	CY4892	induce growth on canavanine	CY4386

In each case an antagonist should cause the test strain to behave more like the control strain.

## B. GTPase $G\alpha_s$ Mutants ( $G\alpha$ Specificity)

test component =  $G\alpha_s$  (Q227L)  
control component =  $G\alpha_s$

Readout	test	effect of $G\alpha_{12}$ -antagonist	control
strain	strain		strain
<i>fus1-His3</i>	CY4880	none	CY4883
<i>fus1-lacZ</i>	CY4886	none	CY4889
<i>fus2-CAN1</i>	CY4895	none	CY4898

In each case a non-specific antagonist would cause the test strain to behave more like the control strain.

Additional media requirements: -TRP for Gα plasmid maintenance in fus1-HIS3 and fus2-CAN1 screens and -TRP -URA for Gα and fus1-lacZ plasmid maintenance in fus1-lacZ screen.

## II. Suppression of Activation by Receptors

Constitutively Activated C5a Receptors  
test component = C5aR\* (P<sub>184</sub>L, activated C5a Receptor)  
control component = C5aR

The C5aR\* mutation has a Leucine residue in place of the Proline residue of the wild-type at position 184 of the amino acid sequence.

Readout	test	strain	effect of Gα <sub>12</sub> antagonist	control
fus1-HIS3	CY4029	inhibit growth of -HIS	+AT (Aminotriazole)	CY2246 strain
fus1-lacZ	CY4901	reduce β-gal activity		CY4904
fus2-CAN1	CY4365	induce growth on	canavanine	CY4362

In each case an antagonist should cause the test strain to behave more like the control strain.

Additional media requirements: -LEU for receptor plasmid maintenance in fus1-HIS3 and fus2-CAN1 screens and -LEU-URA for receptor and fus1-lacZ plasmid maintenance in fus1-lacZ screen, non-buffered yeast media (pH 5.5).

### Example 9: Identification of a surrogate ligand using expression of a random peptide library in yeast expressing an orphan mammalian receptor

FPR1-1 (formyl peptide receptor-like 1) is a structural homolog of the formyl peptide receptor (FPR). FPR is a G protein-coupled receptor, expressed on neutrophils and phagocytic cells, that is stimulated by N-formyl peptides of bacterial origin. Specific binding of the natural ligand, f-Met-Leu-Phe, stimulates transduction of a signal to mobilize calcium,



resulting in cellular changes including chemotaxis and the release of granule contents. Low stringency hybridization of HL60 cDNA libraries with an FPR cDNA probe permitted the identification of the related receptor, FPR-1 (Murphy et al. *supra*). The FPR-1 cDNA encodes a 351 amino acid protein with 69% sequence homology to FPR (Murphy et al. *supra*) FPR and FPR-1 were found to co-localize to human chromosome 19 and to have a tissue expression pattern identical to that of FPR, i.e., expression is restricted to cells of myeloid origin (Murphy et al. *supra*). Ye et al. (*supra*) demonstrated weak binding of f-Met-Leu-Phe (uM concentrations) to fibroblasts transfected with FPR-1 cDNA. In contrast, Murphy et al. (*supra*) could not detect binding of N-formyl peptides to *Xenopus* oocytes transfected with FPR-1 cDNA. FPR-1 appears to be an orphan receptor whose specific ligand differs from the formyl peptide ligands to which FPR responds.

In this example experiments detailing the following will be described: (1) establishment of a strain of yeast designed to express the human orphan G protein-coupled receptor FPR-1; (2) expression of a random peptide library in the aforementioned strain of yeast; and (3) activation of the endogenous yeast pheromone pathway upon stimulation of the FPR-1 receptor by a peptide encoded by a random library expressed within the same strain of yeast.

#### Preparation of FPR-1 Yeast Expression Vector

A plasmid, pFPR-1-L31, containing a 2.6 kb EcoRI-XhoI fragment encoding the FPR-1 cDNA in the Bluescript<sup>TM</sup>SK+ vector was obtained from Philip Murphy (NIH). The sequence encoding FPR-1 was amplified by the polymerase chain reaction using VENT polymerase (New England Biolabs, Inc., Beverly, MA) through 20 cycles and the following oligonucleotide primers:

```
#1 5' GGGCGCCGGTCTCCATGGAAACCACTTCTCCACT
#2 5' GGGCGCCGGTCTCCATGGTGAAGTCACTAGTCTC
```

The PCR product was purified, restricted with BsaI and cloned into Cadus 1651 (pPBX-1), a PGK promoter-driven expression vector, using NcoI and BamHI sites, to yield CADUS 2311. The sequence of the entire insert was determined and found to be identical to the FPR-1 sequence deposited in GenBank (accession number M84562).

#### Preparation of Random Oligonucleotides

##### Library-Recycling Protocol to Identify a Surrogate Ligand

The yeast strain CY1141 (MAT $\alpha$ pha far1\*1441 tbt1-1 fus1-HIS3 can 1 ste14::trp1::LYS2 ste3\*1156 gpa1(41)-Galpha2 lys2 ura3 leu2 trp1 his3) was used in the experiments that follow. CY1141 contains a pheromone inducible HIS3 gene, fus1-HIS3

integrated at the FUS1 locus and a hybrid gene encoding the first 41 amino acids of GPA1 (Yeast G alpha) fused to sequence encoding human G alpha12 (lacking codons encoding the N-terminal 33 amino acids) replacing GPA1 at its chromosomal locus. The yeast STE14 gene is disrupted to lower the basal level of signaling through the pheromone response pathway. The yeast a-factor receptor gene, STE3, is deleted. CY1141 was transformed with Cadus 2311 to yield CY6571, a strain expressing the human orphan receptor, FPR1-1. CY6571 exhibited LIRMA (ligand independent receptor mediated activation), that is, activation of the yeast pheromone pathway in the absence of ligand. It was determined that the yeast growth on selective media that resulted from LIRMA was eliminated by the additional of 2.5 millimolar concentrations of 3-aminotriazole (AT). AT is an inhibitor of the *HIS3* gene product that serves to reduce background growth. Therefore, selection protocols aimed at the identification of surrogate ligands for the FPR1-1 receptor were carried out at this concentration of AT.

CY6571 was inoculated to 10 ml of standard synthetic media (SD) lacking leucine (-Leu) and incubated overnight at 30°C. The 10 ml overnight culture was used to inoculate 50 ml of YEPD; this culture was incubated at 30°C for 4.5-5 hours at which time the cells were harvested and prepared for transformation with DNA encoding a random peptide library [alpha-NNK (6.24.94)] encoding tridecapeptides of random sequence, by electroporation. Post electroporation (in 0.2 cm cuvettes, 0.25  $\mu$ F, 200 $\Omega$ , 1.5 kV) the cells were immediately diluted in 1 ml ice-cold 1M sorbitol and 100  $\mu$ L aliquots were placed onto 10 synthetic media plates (pH6.8) lacking leucine and uracil (-Leu-Ura). The plates were incubated at 30°C for 2-4 days at which time two replicas of each original transformation plate were made to synthetic media (pH6.8) lacking leucine, uracil and histidine and supplemented with 2.5mM AT(-Leu-Ura-His+2.5mM AT). The replicas were incubated at 30°C for 3-5 days. Post incubation the colonies present on the replica sets of two were scraped from the plates into a total of 10 ml of H<sub>2</sub>O (5 ml each plate). The OD<sub>600</sub> of each cell suspension was determined and crude plasmid isolations were done on 8-16 OD units of cells for each pool. A total of eight pools resulted, due to lower numbers of yeast colonies present in four sets of plates. The pellets obtained from these crude plasmid isolations (the so called "smash and grab" technique, Methods in Yeast Genetics - A Laboratory Manual, 1990, M.D. Rose, F. Winston and F. Heiler. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.), were resuspended in 40  $\mu$ L of 10 mM Tris, 1 mM EDTA, pH8.0 and 1  $\mu$ L was used to transform *E. coli* by electroporation (0.1 cm cuvettes, 0.25  $\mu$ F, 200 $\Omega$ , 1.8 kV). Post electroporation the cells were immediately diluted into 1 ml 2XYT media and incubated, with shaking, at 37°C for 30 minutes after which time the cells were used to inoculate 50 ml of 2XYT supplemented with 100  $\mu$ g/ml ampicillin. The 10 resulting cultures were incubated at 37°C overnight. Plasmid DNA was isolated from each of these bacteria cultures using Qiagen columns (Qiagen, Inc., Chatsworth, CA). Each plasmid DNA pellet was

resuspended in 50µL Tris 10mM, EDTA 1 mM, pH 8.0.

Strain CY6571 was transformed with 1µL of each plasmid pool by electroporation. Post electroporation the cells were diluted into 400µL 1M sorbitol. From each electroporated cell suspension, 1µL and 400µL of cells were plated on -Leu-Ura synthetic media, pH6.8 to yield "low density" and "high density" platings. The plates were incubated at 30°C for 3 days, at which time replicas of both the low and high density plates were made to -Leu-Ura-His+2.5mM AT. For those cases where enrichment for a plasmid capable of conferring a His+ phenotype had occurred, this would be reflected by an amplified number of His+ colonies on both the low and high density plates visible at days 2-3, although the amplification would be most obvious on the plates that had received a high density of cells. In the FPR-1 experiment 1/8 pools showed amplification of His+ colonies. The cells were scraped from this plate into 5 mL of H<sub>2</sub>O, the OD<sub>600</sub> of the cell suspension was determined and a crude plasmid isolation was done on 15 OD units of yeast cells. The pellet obtained was resuspended in 40µL 10mM Tris, 1 mM EDTA, pH8.0 and 1µL was used to transform *E. coli*. Plasmid DNA was isolated by miniprep from 3 ml 2XYT cultures of single bacterial colonies resulting from this transformation. 10 DNA pellets (A1 through A10) deriving from individual bacterial colonies were resuspended in 20µL 10 mM Tris 1 mM EDTA, pH8.0 and used to transform CY6571 (containing the FPR-1 expression vector) and CY6263 (CY1141 containing a control expression vector lacking any receptor sequence) by electroporation. Cadus 1625, a control vector lacking sequences encoding a peptide, was included and used to transform both the receptor+ and receptor- strains of yeast. Transformants were first selected on -Leu-Ura, pH6.8 then three yeast transformants of each type (from 11 CY6571 transformations and 11 CY6263 transformations) were patched to -Leu-Ura, pH6.8 to expand the colonies. Once expanded, streaks of the transformants were made on -Leu-Ura-His+2.5mM AT to test for growth in the absence of histidine. All plasmids except the one denoted A2 conferred a growth advantage on media lacking histidine to yeast bearing the FPR-1-encoding plasmid but not to yeast lacking the receptor plasmid. The peptide sequence found to be encoded by plasmids A1 and A3-A10 is: SerLeuLeuTyrLeuThrCysArgProTyrGluAlaMet, and is encoded by the nucleotide sequence 5'-TCT CTG CTT TGG ACT TGT TGT CGG CCT TGG GAG GCG ATG-3'.

# Activation of the Pheromone Response Pathway in Yeast Expressing the FPR-1 Receptor and Peptide Agonist.

For verification of pheromone pathway activation and quantification of the stimulation, the activity of the *fus1* promoter was determined colorimetrically using a *fus1-lacZ* fusion in a parallel set of test strains. CY1141, described above, was used as the recipient strain for these experiments. Transformants contained CADUS 1584 (PRS424-*fus1-lacZ*) in addition to receptor (R+/-) and ligand (L+/-) plasmids. Four strains (bearing the

identical plasmids) were grown overnight in minimal media lacking leucine, uracil, and tryptophan, pH8.6. The overnight cultures were used to inoculate -Leu -Ura -Trp pH8.6 media and these new cultures were grown for approximately 4.5-5 hours to an OD<sub>600</sub> of less than 0.4. Assay of  $\beta$ -galactosidase activity (Guarente 1983) in cells from these cultures yielded the following results:

CY1141/CADUS 2311/peptide A1/CADUS 1584	R+L+	28 units
CY1141/CADUS 2311/CADUS 1625/CADUS 1584	R+L-	3 units
CY1141/CADUS 1289/peptide A1/CADUS 1584	R-L+	3.5 units
CY1141/CADUS 1289/CADUS 1625/CADUS 1584	R-L-	3.9 units

The presence of receptor and peptide-encoding plasmids resulted in an average 8-fold stimulation over background levels of  $\beta$ -galactosidase.

#### **Example 10: Identification of surrogate ligands using expression of a random peptide library in yeast expressing the orphan mammalian receptor, MDR-15.**

In a similar manner a plasmid encoding the monocytic derived receptor monocytic-derived receptor 15 (MDR15; Barella et al. (1995) *Biochem. J.* 309:773-9) was used to construct a yeast strain (CY6573) expressing this receptor. This receptor is an alternative spliced form of the Burkitt's lymphoma receptor 1 (BLR1) encoded by a human Burkitt's lymphoma cDNA (Dobner et al. (1992) *Eur. J. Immunol.* 22, 2795-2799). Strain CY6573 was transformed in a similar manner with the NK13 library, and, following selection on ten -Leu-Ura (4.4 x 10<sup>5</sup> colonies per plate), replica plated to -Leu-Ura-His+ 1mM AT plates. Upon reisolation of plasmid pools and re-transformation into strain CY6573; eight of ten pools showed significantly enriched colony formation on -Leu-Ura-His+ 1mM AT plates. Eight unique plasmids derived from these pools when retransformed into CY6573 conferred growth on -Leu-Ura-His+ 1mM AT plates. One of these plasmids failed to confer growth in a yeast strain lacking the MDR15 receptor.

#### **Example 11: Identification of a ligand using expression of a random peptide library in yeast expressing the human thrombin receptor**

The receptor for thrombin, a G protein-coupled receptor, is present on numerous cell types including platelets, vascular smooth muscle, fibroblasts and on a subset of cells that function in immunity. Thrombin, a serine protease, binds to and cleaves the receptor molecule at residue 41, generating a new receptor N-terminus. The post-cleavage N-terminal residues then act as a "tethered ligand" to activate the receptor molecule (Vu et al. 1994). In platelets, signaling through the thrombin receptor has been shown to result in numerous effects including stimulation of phospholipase C, mobilization of intracellular Ca<sup>2+</sup> and inhibition of adenylyl cyclase.

In this example experiments that detail the following will be described (1) establishment of a strain of yeast designed to express the human G protein-coupled receptor for thrombin; (2) expression of a random peptide library in the afore-mentioned strain of yeast and (3) activation of the endogenous yeast pheromone pathway upon stimulation of the thrombin receptor by peptides encoded by a random library expressed within the same strain of yeast.

#### Preparation of a Yeast Expression Vector for a Mammalian Thrombin Receptor

The human thrombin receptor was amplified by PCR from pCDNA3:Hu-Thr9b-5' (Bristol Myers Squibb) using the following oligonucleotides:

5' GGGCCATGGGGCGCGCGGTTG 3'

5' CCCGGATCCTAAGTTAAGCAGCTTTTGTATAT 3'

The amplified product was purified by gel electrophoresis, restricted with NcoI and BamHI and ligated to NcoI and BamHI-cut CADUS 1871, a PGK promoter-driven expression vector, to yield CADUS 2260. Cloning into CADUS 1871 introduces a novel stop codon preceded by the triplet GlySerVal after the authentic carboxy terminal codon of the human thrombin receptor (threonine). In addition, an invertase signal sequence is fused to the authentic amino terminus of the receptor.

CY7467 exhibited LIRMA (ligand independent receptor mediated activation), that is, activation of the yeast pheromone pathway in the absence of ligand. It was determined that the yeast growth on selective media that resulted from LIRMA was eliminated by the addition of 2.5 millimolar concentrations of 3-aminotriazole (AT). AT is an inhibitor of the HIS3 gene product that serves to reduce background growth. Therefore, selection protocols aimed at the identification of novel peptide ligands for the human thrombin receptor were carried out at this concentration of AT.

#### Preparation of Random Oligonucleotide Library

As described above.

#### Recycling Protocol to Identify a Surrogate Ligand

The yeast strain CY1141 (MAT $\alpha$ pha far1\*1442 tbt1-1 fus1-HIS3 can1 ste14::tp1::LYS2 ste3\*1156 gpa1(41)-Galpha12 lys2 ura3 leu2 tp1 his3) was transformed with CADUS 2260 to yield strain CY7467, expressing the human thrombin receptor. CY7467 was inoculated to 10 mls of standard synthetic media (SD) lacking leucine (-Leu) and incubated overnight at 30 C. The 10 ml overnight culture was used to inoculate 50 mls

of YEPD media; this culture was incubated at 30 °C for 4.5-5 hours at which time the cells were harvested and prepared for transformation with DNA encoding a random peptide library [alpha-NNK (6.24.94)] by electroporation. Post electroporation (in 0.2 cm cuvettes, 0.25 mF, 200W, 1.5 kV) the cells were immediately diluted in 1 ml ice-cold 1M sorbitol and 100mL aliquots were plated onto 10 synthetic media plates (pH6.8) lacking leucine and uracil (-Leu-Ura). The plates were incubated at 30 °C for 2-4 days at which time two replicas of each original transformation plate were made to synthetic media (pH6.8) lacking leucine, uracil and histidine and supplemented with 2.5mM AT(-Leu-Ura-His+ 2.5mM AT). The replicas were incubated at 30 °C for 3-5 days. Post incubation the colonies present on the replica sets of two were scraped from the plates into a total of 10 mls of H<sub>2</sub>O (5 mls each plate). The OD<sub>600</sub> of each cell suspension was determined and crude plasmid isolations were done on 8-16 OD units of cells for each pool. A total of ten pools resulted. The pellets obtained from these crude plasmid isolations were resuspended in 40mL of 10 mM Tris, 1 mM EDTA, pH8.0 and 1mL was used to transform *E. coli* by electroporation (0.1 cm cuvettes, 0.25 mF, 200W, 1.8 kV). Post electroporation the cells were immediately diluted into 1 ml 2XYT media and incubated, with shaking, at 37 °C for 30 minutes after which time the cells were used to inoculate 50 mls of 2xYT supplemented with 100 ug/ml ampicillin. The 10 resulting cultures were incubated at 37 °C overnight. Plasmid DNA was isolated from each of these bacterial cultures using Qiagen columns (Qiagen, Inc., Chatsworth, CA). Each plasmid DNA pellet was resuspended in 50mL Tris 10mM, EDTA 1 mM, pH 8.0.

Strain CY7467 was transformed with 1mL of each plasmid pool by electroporation. Post electroporation the cells were diluted into 400mL 1M sorbitol. From each electroporated cell suspension, 1mL and 400mL of cells were plated on -Leu-Ura synthetic media, pH6.8 to yield "low density" and "high density" platings. The plates were incubated at 30 °C for 3 days, at which time replicas of both the low and high density plates were made to -Leu-Ura-His+ 2.5mM AT. For those cases where enrichment for a plasmid capable of conferring a His+ phenotype had occurred, this would be reflected by an amplified number of His+ colonies on both the low and high density plates visible at days 2-3, although the amplification would be most obvious on the plates that had received a high density of cells. In this experiment 3/10 pools showed amplification of His+ colonies. The cells from each of these plates were scraped into 5 mls of H<sub>2</sub>O, the OD<sub>600</sub> of the cell suspensions were determined and crude plasmid isolations were done on 8-16 OD units of yeast cells. The pellets obtained were resuspended in 40mL 10 mM Tris, 1 mM EDTA, pH8.0 and 1mL was used to transform *E. coli*. Plasmid DNA was isolated by miniprep from 3 ml 2XYT cultures of single bacterial colonies resulting from these transformations (three bacterial colonies for each DNA pool were processed in this way). DNAs deriving from three individual bacterial colonies per pool were resuspended in 20mL 10 mM Tris 1 mM EDTA, pH8.0. The three DNAs derived per pool were sequenced and found to encode identical peptides. Thus three

differing DNA sequences were derived, one representing each amplified pool. One plasmid representing each of the three original amplified pools was used to transform CY7467 (containing the thrombin receptor expression vector) and CY6263 (CY1141 containing a control expression vector lacking any receptor sequence) by electroporation. CADUS 1625, a control vector lacking sequences encoding a peptide was included and used to transform both the receptor+ and receptor- strains of yeast. CADUS 1651, a control vector lacking sequences encoding a receptor included and used to transform both the ligand+ and ligand- strains of yeast. Transformants were first selected on -Leu-Ura, pH6.8, then two yeast transformants of each type were patched to -Leu-Ura, pH6.8 to expand the colonies. Once expanded, streaks of the transformants were made on -Leu-Ura-His+ 2.5mM AT to test for growth in the absence of histidine. One of the three plasmids tested conferred a growth advantage on media lacking histidine to yeast bearing the thrombin-encoding plasmid but not to yeast lacking the receptor plasmid. The peptide sequence encoded by this plasmid is: Val-Cys-Pro-Ala-Arg-Tyr-Val-Leu-Pro-Gly-Pro-Val-Leu and was encoded by the nucleotide sequence GTT TGT CCT GCG CGT TAT GTG CTG CCT GCG CCT GTT TTG.

Table 1. Detection of CSA production in Yeast by ELISA.

	R-L-	R-L-	R-L+	R+L+
[C5a] in culture	n.d.	n.d.	0.64 ng/ml	0.5 ng/ml
[C5a] released	n.d.	n.d.	0.8 ng/ml	0.6 ng/ml
from lysed cells*			0.97 nM	0.73 nM

10 C5a was detected by enzyme-linked immunosorbent assay (ELISA). Molar concentrations were calculated using MW=8273 as predicted by C5a sequence.

\* Determined by pelleting cells, resuspending cells in the original volume, breaking yeast with glass beads and assaying the resulting supernatant.

n.d.=not done

enop 700= .p .u



Table 2. Coupling of the Gsa receptor to Gα chimeras in yeast.

Chimera	Expression	Result
GPA1 <sup>+</sup> -Gα12	single copy, integrated, GPA1 promoter	Good signal to noise ratio: efficient coupling to yeast G $\gamma$ .
	single copy, integrated, GPA1 promoter	Poor signal to noise ratio: high background due to poor coupling to yeast G $\gamma$ , high LIRMA*.
GPA1 <sup>+</sup> -Gα13	single copy, integrated, GPA1 promoter	coupling to yeast G $\gamma$ , high LIRMA*.
	single copy, integrated, GPA1 promoter	Unacceptably high background due to poor coupling to
GPA1 <sup>+</sup> -Gα12	low copy plasmid, GPA1 promoter	Signal equal to that with GPA1 <sup>+</sup> -Gα12, however, background is greater.
	low copy plasmid, GPA1 promoter	Poor signal to noise ratio, high background due to poor coupling to yeast G $\gamma$ , high LIRMA*.
GPA1 <sup>+</sup> -Gα16	low copy plasmid, GPA1 promoter	coupling to yeast G $\gamma$ , high LIRMA*.
	low copy plasmid, GPA1 promoter	Unacceptably high background due to poor coupling to
GPA1 <sup>+</sup> -Gαs	low copy plasmid, GPA1 promoter	Yeast G $\gamma$ , high LIRMA*.
	low copy plasmid, GPA1 promoter	Unacceptably high background due to poor coupling to

\*LIRMA = Ligand Independent Receptor Mediated Activation. With this phenomenon, there is an increase in growth on selective media for strains containing heterologous receptor in the absence of ligand. It is possible that some receptor antagonists would decrease LIRMA. It has been noted (Milano, et al. 1994) that specific antagonists reduce LIRMA of the  $\beta$ 2 adrenergic receptor when that receptor is overexpressed in transgenic mice. LIRMA may be exploited in several ways, including the identification of antagonists capable of reducing the phenomenon. A subset of antagonists would be expected to affect the receptor conformation in such a way as to prevent the downstream signalling that occurs in the absence of agonist. LIRMA can be exploited to identify new G protein-coupled receptors by expressing cDNA clones in yeast strains expressing those chimeric G proteins which couple only poorly to yeast G $\gamma$ . In addition, LIRMA may permit the identification of inhibitors that are specific for G proteins.

Table 3. Coupling of  $\alpha$  switch region hybrids to the pheromone response pathway.

Protein	GPA1 amino acid sequences	Gas amino acid sequences	phenotype
GPA1	1-472	none	Couples with $\alpha$ hy
Gas	none	1-394	Couples with $\alpha$ hy weakly
GPA1-S	1-41	42-394	Couples with $\alpha$ hy weakly
SGS	297-333	1-201 + 237-394	Does not couple with $\alpha$ hy
GPA1-SGS	1-41 + 297-333	42-201 + 237-394	Couples with $\alpha$ hy weakly

We claim:

126  
- 95 -

1. A mixture of recombinant cells, each cell of which comprises:
  - (i) an expressible recombinant gene encoding a heterologous receptor protein whose signal transduction activity is modulated by interaction with an extracellular signal; and
  - (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide, wherein collectively the mixture of cells expresses a variegated population of said receptor effector polypeptides, and modulation of the signal transduction activity of the receptor protein by a test polypeptide provides a detectable signal.
2. A mixture of recombinant cells, each cell of which comprises:
  - (i) a heterologous receptor protein whose signal transduction activity is modulated by interaction with an extracellular signal;
  - (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide; and
  - (iii) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the receptor protein, wherein collectively the mixture of cells expresses a variegated population of test polypeptides as receptor effectors.
3. The cells of claim 2, wherein the receptor is a nuclear receptor.
4. The cells of claim 2, wherein the receptor is a cell surface receptor.
5. A mixture of recombinant cells, each cell of which comprises:
  - (i) a receptor protein whose signal transduction activity is modulated by interaction with an extracellular signal;
  - (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide; and
  - (iii) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the receptor protein, wherein collectively the mixture of cells expresses a variegated population of test polypeptides as receptor effectors.

6. The cells of claim 5, wherein the receptor is a nuclear receptor.
7. The cells of claim 5, wherein the receptor is a cell surface receptor.
8. A mixture of recombinant cells, each cell of which comprises:
  - (i) a cell surface receptor protein whose signal transduction activity is modulated by interaction with an extracellular signal; and
  - (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide including a signal sequence for secretion, wherein collectively the mixture of cells expresses a variegated population of test polypeptides as receptor effectors, and modulation of the signal transduction activity of the receptor protein by a test polypeptide provides a detectable signal.
9. The recombinant cells of claim 8, wherein each cell further comprises a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the cell surface receptor protein, expression of the reporter gene providing the detectable signal.
10. The recombinant cells of claim 8, wherein the reporter gene encodes a gene product that gives rise to a detectable signal selected from the group consisting of: color, fluorescence, luminescence, cell viability, relief of a cell nutritional requirement, cell growth, and drug resistance.
11. The recombinant cells of claim 9, wherein the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase, beta-galactosidase and secreted alkaline phosphatase.
12. The recombinant cells of claim 9, wherein the reporter gene encodes a gene product which confers a growth signal.
13. The recombinant cells of claim 9, wherein the reporter gene encodes a gene product for growth in media containing aminotriazole or canavanine.
14. The recombinant cells of claim 8, wherein the detectable signal comprises intracellular calcium mobilization.
15. The recombinant cells of claim 8, wherein the detectable signal comprises a significant change in intracellular protein phosphorylation.

BB  
B5

BB  
B4

BB  
B3

16. The recombinant cells of claim 8, wherein the detectable signal comprises increases in phospholipid metabolism.

17. The recombinant cells of claim 8, wherein each cell further comprises a heterologous gene construct encoding the receptor protein.

18. The recombinant cells of claim 8, wherein the receptor protein is a G-protein coupled receptor.

19. The recombinant cells of claim 18, wherein the G-protein coupled receptor is selected from the group consisting of: a chemotactant peptide receptor, a neuropeptide receptor, a light receptor, a neurotransmitter receptor, a cyclic AMP receptor, and a polypeptide hormone receptor.

20. The recombinant cells of claim 8 wherein the receptor protein is a receptor tyrosine kinase.

21. The recombinant cells of claim 20, wherein the receptor tyrosine kinase is an EPH receptor.

22. The recombinant cells of claim 8, wherein the receptor protein is an orphan receptor.

23. The recombinant cells of claim 8, which recombinant cells are yeast cells.

24. The recombinant cells of claim 8, which recombinant cells are mammalian cells.

25. The recombinant cells of claim 8, wherein the variegated population of test polypeptides includes at least  $10^3$  different test polypeptides.

26. A recombinant cell, comprising:

- (i) an expressible recombinant gene encoding a heterologous cell surface receptor protein whose signal transduction activity is modulated by extracellular signals;
- (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide including a signal sequence for secretion; and
- (iii) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal

transduction activity of the cell surface receptor protein.  
The recombinant cell of claim 26, wherein the reporter gene encodes a gene product that gives rise to a detectable signal selected from the group consisting of: color, fluorescence, luminescence, cell viability relief of a cell nutritional requirement, cell growth, and drug resistance.

28. The recombinant cell of claim 26, wherein the receptor protein is a G-protein coupled receptor.  
29. The recombinant cell of claim 28, wherein the G-protein coupled receptor is selected from the group consisting of: a chemotactant peptide receptor, a neuropeptide receptor, a light receptor, a neurotransmitter receptor, a cyclic AMP receptor, and a polypeptide hormone receptor.

30. The recombinant cell of claim 26, wherein the receptor protein is a receptor tyrosine kinase.  
31. The recombinant cell of claim 30, wherein the receptor tyrosine kinase is an EPH receptor.

32. The recombinant cell of claim 26, wherein the receptor protein is an orphan receptor.

33. The recombinant cell of claim 26, wherein the receptor protein is a cytokine receptor.

34. The recombinant cell of claim 26, wherein the receptor protein is an MIRK.

35. The recombinant cell of claim 26, which recombinant cell is a yeast cell.

36. The recombinant cell of claim 35, which yeast cells is a *Saccharomyces* cell.

37. The recombinant cell of claim 35, which yeast cells is a *Schizosaccharomyces* cell.

38. The recombinant cell of claim 26, which cells are mammalian cells.

39. A mixture of recombinant cells, each cell of which comprises:

85  
B7C/H

- (i) an expressible recombinant gene encoding a heterologous cell surface receptor protein whose signal transduction activity is modulated by extracellular signals;
  - (ii) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide including a signal sequence for secretion; and
  - (iii) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the cell surface receptor protein,
- wherein collectively the mixture of cells expresses a variegated population of test polypeptides.

40. The recombinant cells of claim 39, wherein the receptor protein is a G-protein coupled receptor.

coupled receptor.

41. The recombinant cells of claim 40, wherein the G-protein coupled receptor is selected from the group consisting of: a chemoattractant peptide receptor, a neuropeptide receptor, a light receptor, a neurotransmitter receptor, a cyclic AMP receptor, and a polypeptide hormone receptor.

42. The recombinant cell of claim 40, wherein the G-protein coupled receptor is selected from the group consisting of:  $\alpha 1$ A-adrenergic receptor,  $\alpha 1$ B-adrenergic receptor,  $\alpha 2$ -

adrenergic receptor,  $\alpha_2$ B-adrenergic receptor,  $\beta_1$ -adrenergic receptor,  $\beta_2$ -adrenergic receptor,  $\beta_3$ -adrenergic receptor, m1 acetylcholine receptor (AChR), m2 AChR, m3 AChR, m4 AChR, m5 AChR, D1 dopamine receptor, D2 dopamine receptor, D3 dopamine receptor, D4 dopamine receptor, D5 dopamine receptor, A1 adenosine receptor, A2b adenosine

(neurokinin A), EMIP receptor, MLP-like receptor, angiotensin II type I, endothelin ETA, receptor, 5-HIA, 5-HIB, CHII-like, 5-HID, CHID-like, substance K  
endothelin ETB, thrombin, growth hormone-releasing hormone (GHRH), vasoactive  
intestinal peptide, oxytocin, somatostatin SSTRI and SSTR2, SSTR3, cannabimoid, follicle

stimulating hormone (TSH), thyrotropin (T/H/C), thyroid stimulating hormone (TSH), thromboxane A<sub>2</sub>, platelet-activating factor (PAF), C5a anaphylatoxin, Interleukin 8 (IL-8) IL-8RA, IL-8RB, Delta Opioid, Kappa Opioid, mip-1/RANTES, Rhodopsin, Red opsin, Green opsin, Blue opsin, metabotropic glutamate mGluR1-6, histamine H<sub>2</sub>, ATP, neuropeptide Y, amyloid protein precursor, insulin-like growth factor II, bradykinin, gonadotropin-releasing hormone, cholecystokinin, melanocyte stimulating hormone receptor, antidiuretic hormone receptor, glucagon receptor, and adrenocorticotrophic hormone II.

43. The recombinant cells of claim 39, wherein the receptor protein is a receptor tyrosine kinase.

tyrosine kinase.

44. The recombinant cells of claim 43, wherein the receptor tyrosine kinase is an EPH receptor.

45. The yeast cell of claim 44, wherein the receptor is selected from the group consisting of: *eph*, *ek*, *ekk*, *sek*, *mek4*, *hek*, *hek2*, *ekk*, *erk*, *tyro1*, *tyro4*, *tyro5*, *tyro6*, *tyro11*, *cek4*, *cek5*, *cek6*, *cek7*, *cek8*, *cek9*, *cek10*, *bsk*, *rik1*, *rik2*, *rik3*, *myk1*, *myk2*, *ehk1*, *ehk2*, *pagliaccio*, *hik*, *erk* and *nuk* receptors.

46. The recombinant cell of claim 39, wherein the receptor protein is a cytokine receptor.

47. The recombinant cell of claim 39, wherein the receptor protein is an MMR receptor.

48. The recombinant cell of claim 39, wherein the receptor protein is an orphan receptor.

49. The recombinant cell of claim 39, which recombinant cell is a yeast cell.

50. The recombinant cell of claim 49, which yeast cells is a *Saccharomyces* cell.

51. The recombinant cell of claim 49, which yeast cells is a *Schizosaccharomyces* cell.

52. The recombinant cell of claim 39, which cells are mammalian cells.

53. The recombinant cells of claim 39, wherein the variegated population of test polypeptides includes at least  $10^3$  different test polypeptides.

54. A method for identifying potential receptor effectors comprising:

- (i) providing a mixture of recombinant cells, each cell of which comprises a receptor protein whose signal transduction activity is modulated by interaction with an extracellular signal, and
- (b) an expressible recombinant gene encoding a heterologous test polypeptide wherein the mixture of cells collectively express a variegated population of test polypeptides, and modulation of the signal transduction activity of the receptor protein by a test polypeptide provides a detection signal; and
- (ii) isolating cells from the mixture which exhibit the detection signal.



55. The method of claim 54, wherein the cell receptor is a cell surface receptor.  
56. The method of claim 55, wherein the heterologous test polypeptide includes a signal sequence for secretion.

57. The method of claim 54, wherein each cell of the mixture further comprises a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the cell surface receptor protein, expression of the reporter gene providing the detection signal.

58. The method of claim 57, wherein the reporter gene encodes a gene product that gives rise to a detection signal selected from the group consisting of: color, fluorescence, luminescence, cell viability relief of a cell nutritional requirement, cell growth, and drug resistance.

59. The method of claim 58, wherein the reporter gene encodes a gene product selected from the group consisting of chloramphenicol acetyl transferase, beta-galactosidase and secreted alkaline phosphatase.

60. The method of claim 58, wherein the reporter gene encodes a gene product which confers a growth signal.

61. The method of claim 58, wherein the reporter gene encodes a gene product for growth in media containing aminotriazole or canavanine.

62. The method of claim 54, wherein the detection signal comprises intracellular calcium mobilization.

63. The method of claim 54, wherein the detection signal comprises a statistically significant change in intracellular protein phosphorylation.

64. The method of claim 54, wherein the detection signal comprises changes in phospholipid metabolism.

65. The method of claim 54, wherein each cell of the mixture further comprises a heterologous gene construct encoding the receptor protein.

66. The method of claim 54, wherein the receptor protein is a G-protein coupled receptor.

67. The method of claim 66, wherein the G-protein coupled receptor is selected from the group consisting of: a chemoattractant peptide receptor, a neuropeptide receptor, a light receptor, a neurotransmitter receptor, a cyclic AMP receptor, and a polypeptide hormone receptor.

68. The method of claim 54, wherein the receptor protein is a receptor tyrosine kinase.

69. The method of claim 68, wherein the receptor tyrosine kinase is an EPH receptor.

70. The method of claim 54, wherein the receptor protein is a cytokine receptor.

71. The method of claim 54, wherein the receptor protein is an orphan receptor.

72. The method of claim 54, wherein the recombinant cells are yeast cells.

73. The method of claim 54, which recombinant cells are mammalian cells.

74. The method of claim 54, wherein the variegated population of test polypeptides includes at least 10<sup>3</sup> different test polypeptides.

75. A method for identifying effectors of a cell surface receptor comprising:  
 (i) providing a mixture of recombinant cells, each cell of which comprises an expressible recombinant gene encoding a heterologous cell surface receptor protein whose signal transduction activity is modulated by extracellular signals,  
 (b) an expressible recombinant gene encoding a heterologous potential receptor effector polypeptide including a signal sequence for secretion, and  
 (c) a reporter gene construct containing a reporter gene in operative linkage with one or more transcriptional regulatory elements responsive to the signal transduction activity of the cell surface receptor protein, wherein the mixture of cells collectively express a variegated population of test polypeptides, and modulation of the signal transduction activity of the receptor protein by a test polypeptide causes a statistically significant change in the level of expression of the reporter gene; and

506  
B11

76. A method for identifying ligands for an orphan cell surface receptor comprising:
- (i) providing a mixture of recombinant cells, each cell of which comprises  
(a) a heterologous gene encoding an orphan cell surface receptor whose  
signal transduction activity is modulated by extracellular signals; and  
(b) an expressible recombinant gene encoding a heterologous test  
polypeptide including a signal sequence for secretion,  
wherein the mixture of cells collectively express a variegated population of test  
polypeptides, and modulation of the signal transduction activity of the orphan  
receptor protein by a test polypeptide provides a detection signal; and
  - (ii) isolating cells from the mixture which exhibit the detection signal.

Add  
B13

The present invention makes available a rapid, effective assay for screening and identifying pharmacologically effective compounds that specifically interact with and modulate the activity of a cellular receptor or ion channel. The subject assay enables rapid screening of large numbers of polypeptides in a library to identifying those polypeptides which induce or antagonize receptor bioactivity. The subject assay is particularly amenable for identifying surrogate ligands for orphan receptors.

*Abstract*



135  
-104-

08/587895

Attorney's

Docket

Number CPI-017

Declaration, Petition and Power of Attorney For Patent Application.

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name,

I believe I am the original, first and sole inventor of the subject matter which is claimed and for which a patent is sought on the invention entitled

**"METHODS AND COMPOSITIONS FOR IDENTIFYING RECEPTOR EFFECTORS"**

the specification of which is filed herewith in the U.S. Patent and Trademark Office.

I do not know and do not believe that the subject matter of this application was ever known or used in the United States before my invention thereof or patented or described in any printed publication in any country before my invention thereof or more than one year prior to the date of this application, and that said subject matter has not been patented or made the subject of an issued inventor's certificate in any country foreign to the United States on an application filed by me or my legal representatives or assigns more than twelve months prior to the date of this application; that I acknowledge my duty to disclose information of which I am aware which is material to the examination of this application, that no application for patent or inventor's certificate on the subject matter of this application has been filed by me or my representatives or assigns in any country foreign to the United States, except those identified below, and that I have reviewed and understand the contents of the specification, including the claims as amended by any amendment referred to herein.

I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, §1.56(a).

PRIORITY CLAIM

I hereby claim foreign priority benefits under Title 35, United States Code, §119 of any foreign application(s) for patent or inventor's certificate or of any PCT international application(s) designating at least one country other than the United States of America listed below and have also identified below any foreign application(s) for patent or inventor's certificate or any PCT international application(s) designating at least one country other than the United States of America having a filing date before that of the application(s) of which priority is claimed.

Check one:

☒ no such applications have been filed.

☐ such applications have been filed as follows

EARLIEST FOREIGN APPLICATION(S), IF ANY FILED WITHIN 12 MONTHS  
(6 MONTHS FOR DESIGN) PRIOR TO THIS U.S. APPLICATION

Country	Application Number	Date of Filing (month,day,year)	Priority Claimed Under 35 USC 119
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No
			<input type="checkbox"/> Yes <input type="checkbox"/> No

ALL FOREIGN APPLICATION(S), IF ANY, FILED MORE THAN 12 MONTHS  
(6 MONTHS FOR DESIGN) PRIOR TO THIS U.S. APPLICATION


# CLAIM FOR BENEFIT OF EARLIER U.S./PCT APPLICATION(S)

I hereby claim the benefit under Title 35, United States Code, §120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, §112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, §1.56(a) which occurred between the filing date of the prior application and the national or PCT international filing date of this application. As to subject matter of this application which is common to my earlier United States application, if any, described below, I do not believe that the same was ever known or used in the United States before my invention thereof or patented or described in any printed publication in any country before my invention thereof or more than one year prior to said earlier application, or in public use or on sale in the United States more than one year prior to said earlier application, that the said common subject matter has not been patented or made the subject of an inventor's certificate issued before the date of said earlier application in any country foreign to the United States on an application, filed by me or my legal representatives or assigns more than twelve months prior to said application and that no application for patent or inventor's certificate on said subject matter has been filed by me or my representatives or assigns in any country foreign to the United States except those identified herein.

<u>(Application Serial No.)</u>	<u>(Filing Date)</u>	<u>(Status)</u>
<u>(Application Serial No.)</u>	<u>(Filing Date)</u>	<u>(Status)</u>
<u>(Application Serial No.)</u>	<u>(Filing Date)</u>	<u>(Status)</u>

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorneys and/or agents to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

John A. Labive, Jr.	Reg. N. 19,788	Jeremiah Lynch	Reg. No. 17,425
W. Hugo Liepmann	Reg. No. 20,407	Amy E. Mandragouras	Reg. No. 36,207
James E. Cockfield	Reg. No. 19,162	Elizabeth A. Hanley	Reg. No. 33,505
Thomas V. Smurzynski	Reg. No. 24,798	Matthew P. Vincent	Reg. No. 36,709
Ralph A. Loren	Reg. No. 29,325	Paul Louis Myers	Reg. No. 35,965
Thomas J. Engellenner	Reg. No. 28,711	Beth E. Arnold	Reg. No. 35,430
William C. Geary III	Reg. No. 31,359	Anthony A. Laureniano	Reg. No. 38,220
Giulio A. DeConti, Jr.	Reg. No. 31,503	Jane E. Remillard	Reg. No. 38,872
Michael I. Falkoff	Reg. No. 30,833	Jean M. Silveri	Reg. No. 39,030
Ann Lamport Hammitte	Reg. No. 34,858	Mark A. Kurtsko	Reg. No. 38,944
John V. Bianco	Reg. No. 36,748	Edward J. Kelly	Reg. No. 38,936

Send Correspondence to: Labive & Cockfield, 60 State Street, Boston, MA 02109

Direct Telephone Calls to: Matthew P. Vincent, (617) 227-7400

Wherefore I petition that letters patent be granted to me for the invention or discovery described and claimed in the attached specification and claims, and hereby subscribe my name to said specification and claims and to the foregoing declaration, power of attorney, and this petition.

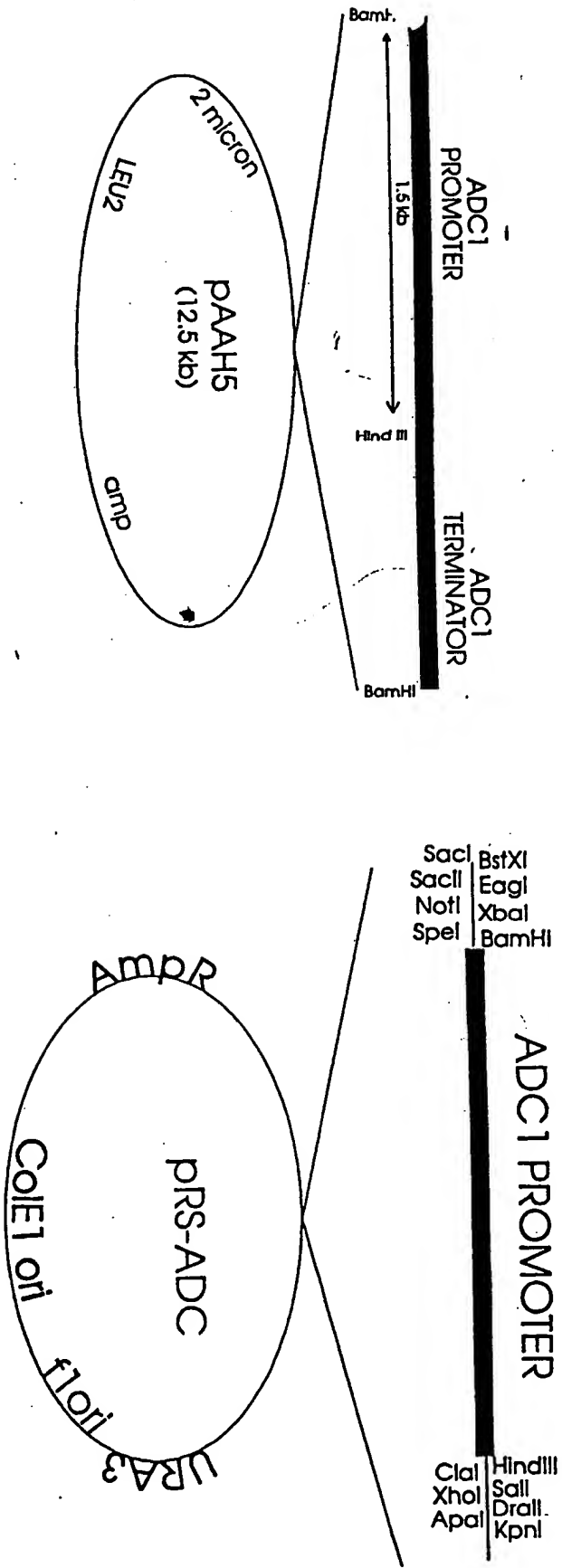
I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Full name of first inventor	Christine A. Klein
Inventor's signature	Date
Residence	Teatown Lake Reservation, Spring Valley Road, Ossining, New York 10562
Citizenship	USA



Full name of second inventor		Andrew J. M. Murphy	
Inventor's signature		Date	
Residence		17 Windsor Place, Montclair, New Jersey 07043	
Citizenship		USA	

FIGURE 1



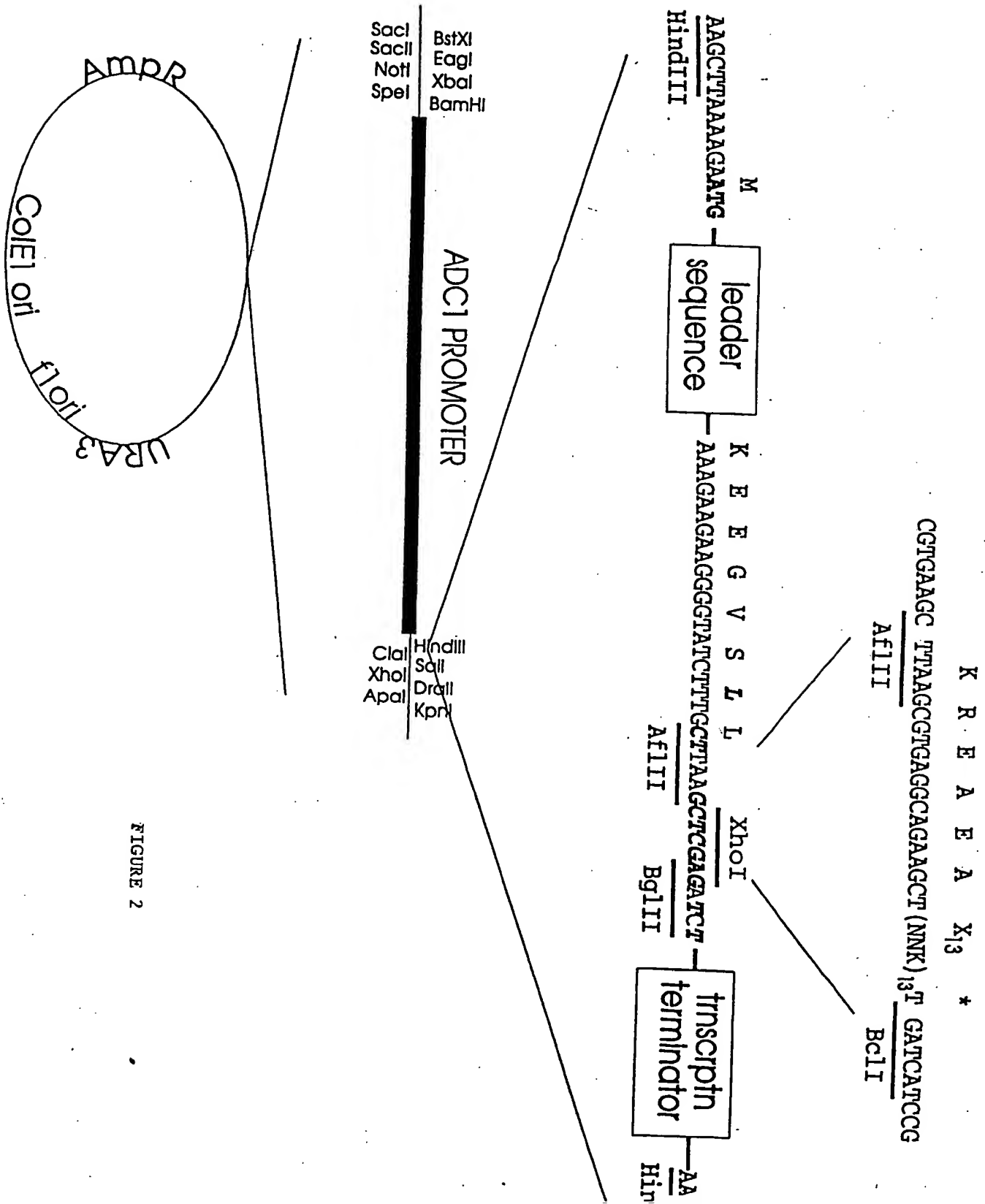


FIGURE 2

S S E K K D N X<sub>11</sub> C V I A \*

GGTAC TCGAGTCAAAGAAGACAAAC (NNK)<sub>11</sub> TGTGTTATGCTTAA GTACG

XhoI

AflIII

SacI

XhoI AflIII

AGCTTTCGATAGAAATG  
HindIII M

MfaI  
leader

GCCGCTCCAAGAAGAAGACC TCGAGCTCGCTTAAG  
A A P K E K T S S

transcript  
terminator

AAG  
Hinc

ADC1 PROMOTER

BstXI  
EagI  
XbaI  
BamHI  
SacI  
SacII  
NotI  
SpeI

HindIII  
SalI  
DraIII  
KpnI  
ClaI  
XhoI  
ApaI

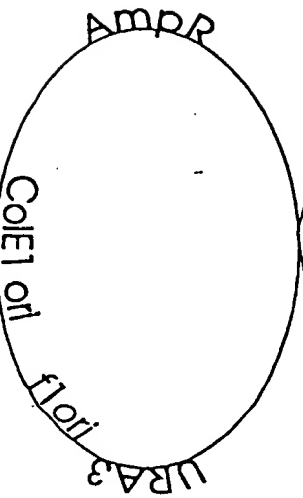


FIGURE 3

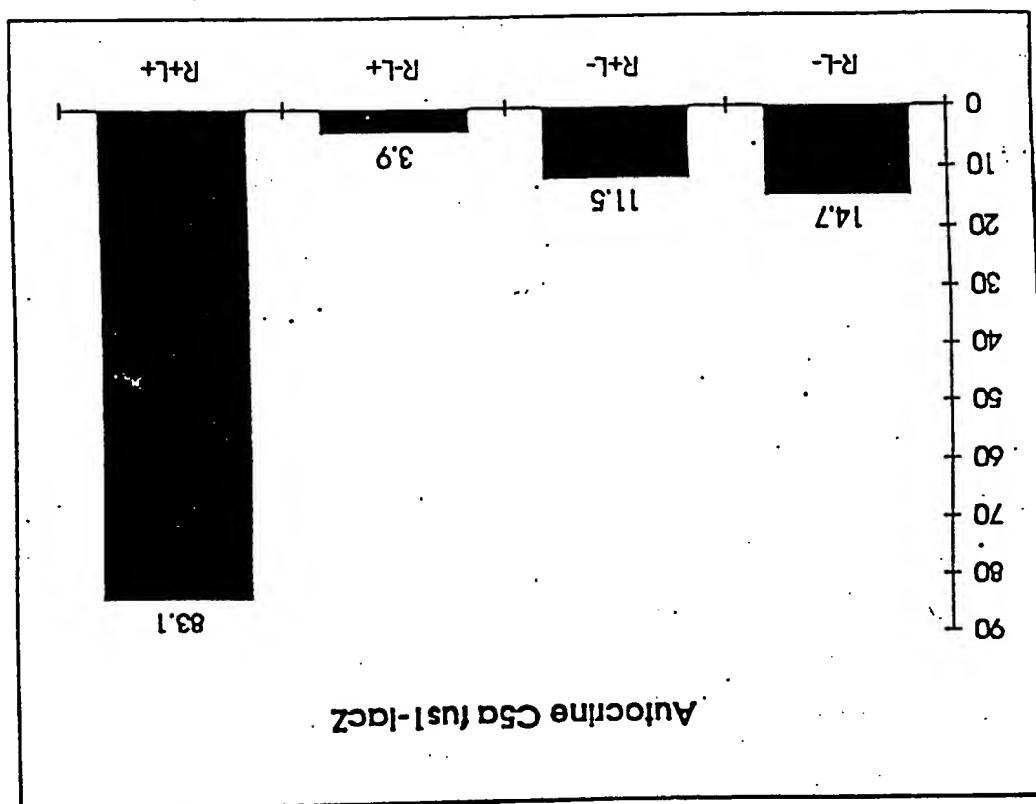


FIGURE 4

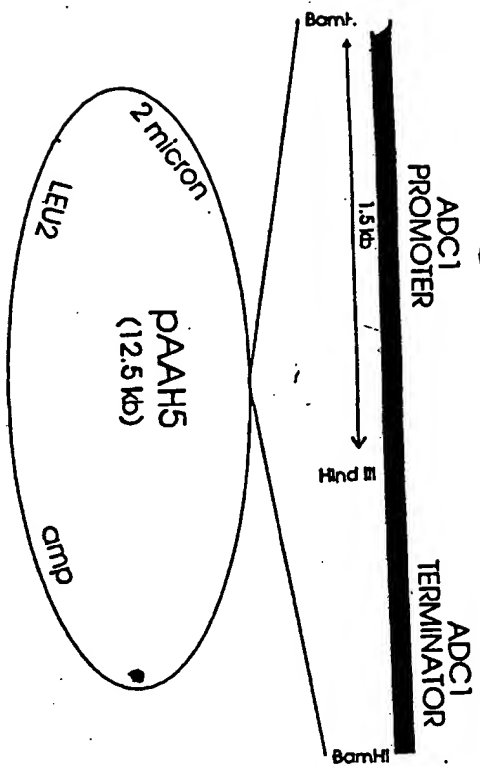
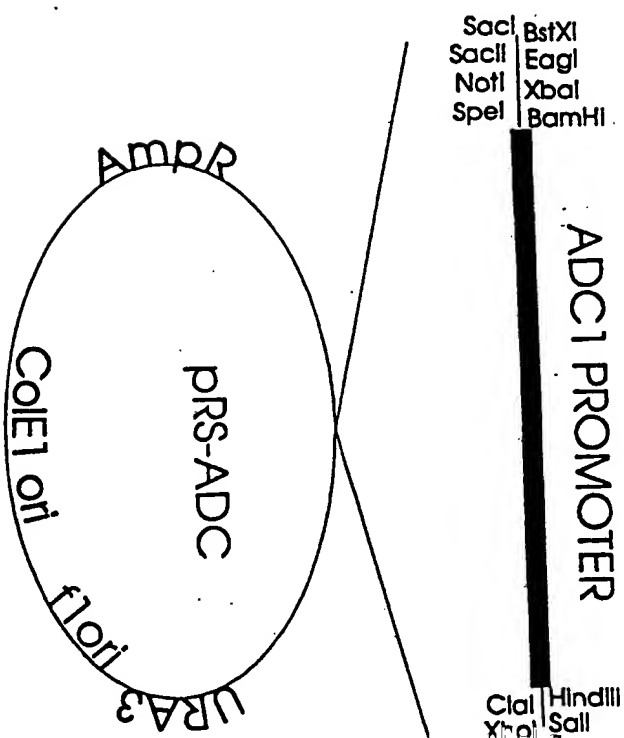
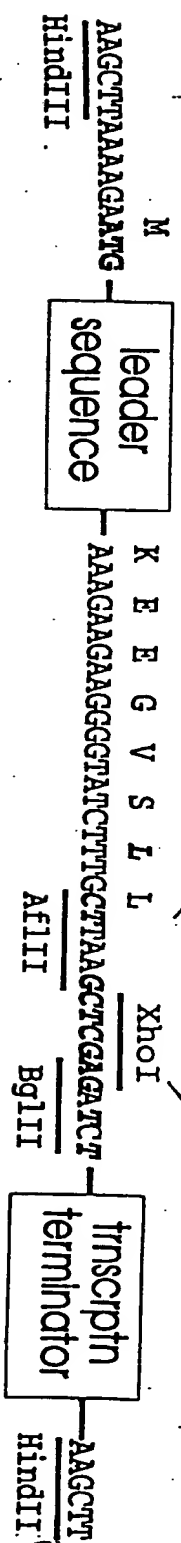


FIGURE 1



K R E A E A X<sub>13</sub> \*

CGTGAAGC TTAAGCGTGAGGAGAAAGCT(NNK)<sub>13</sub>T GATCATCCG  
AflII BclI



ADC1 PROMOTER

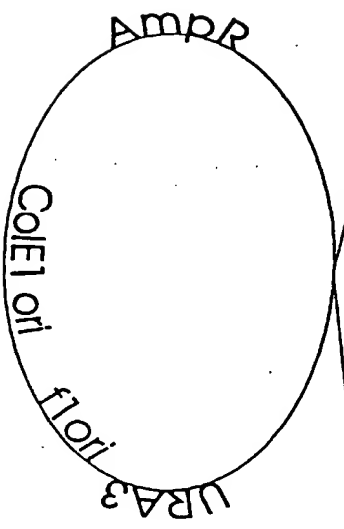


FIGURE 2

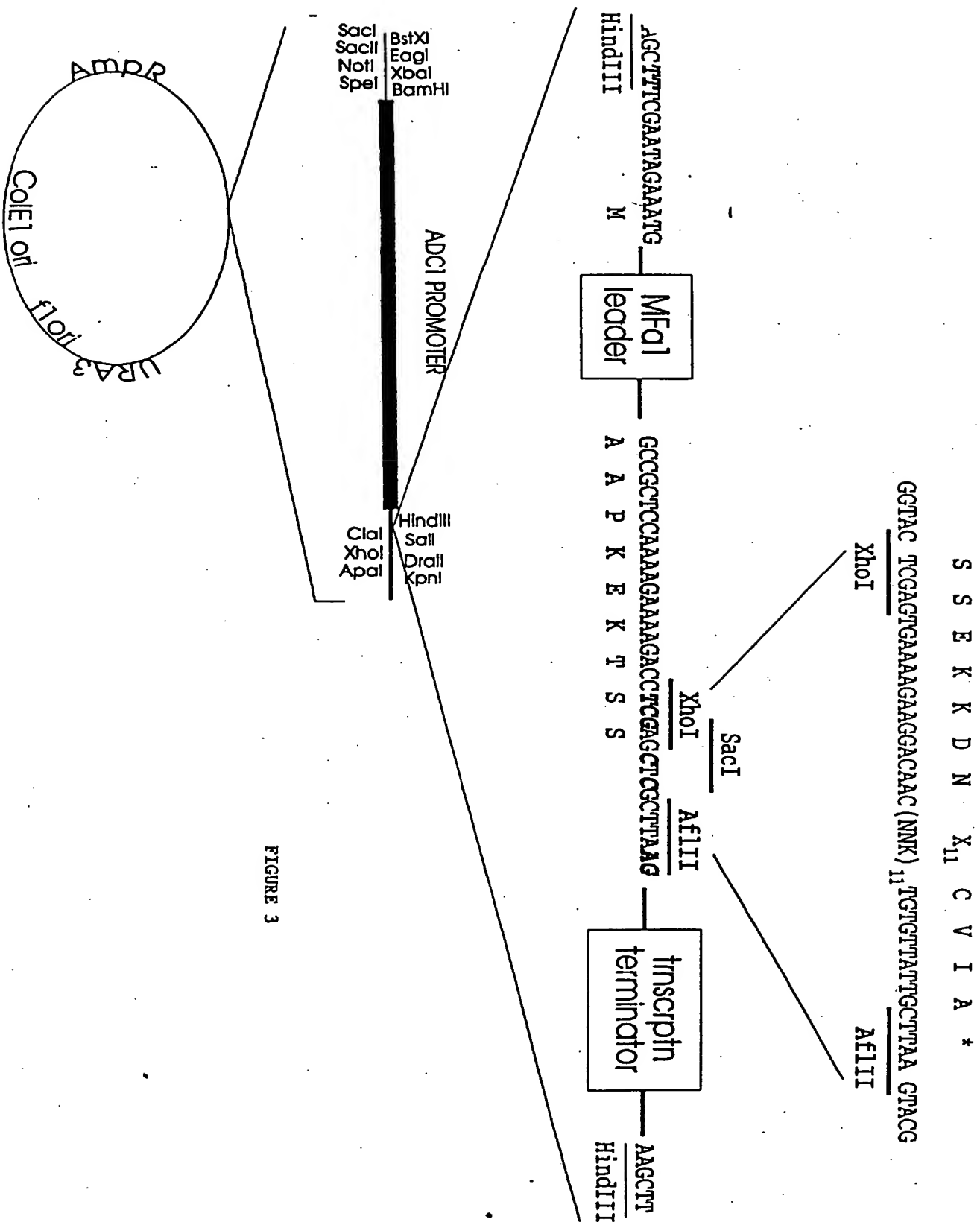


FIGURE 3



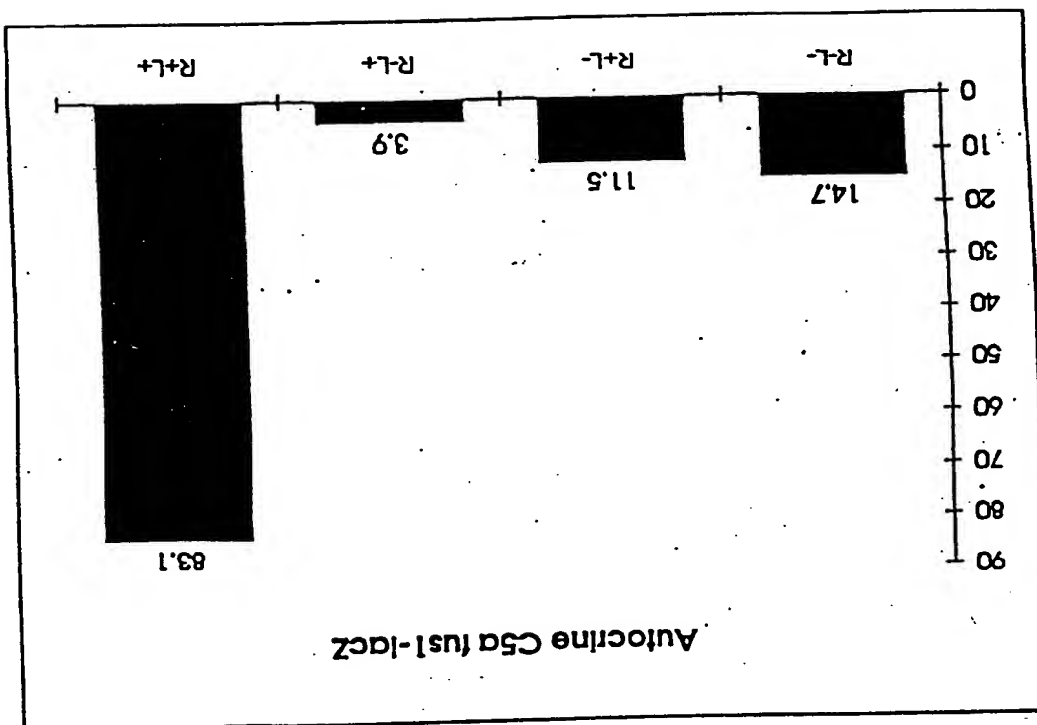
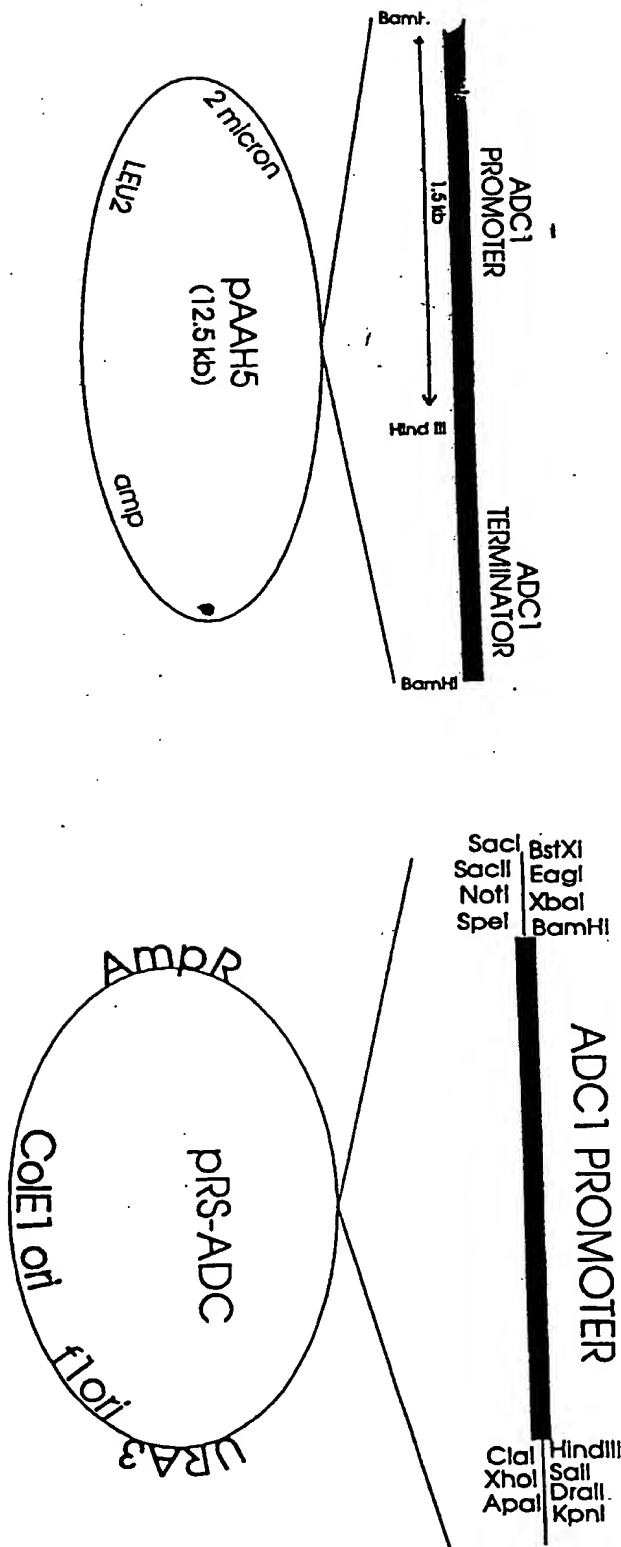


FIGURE 4

08/587895

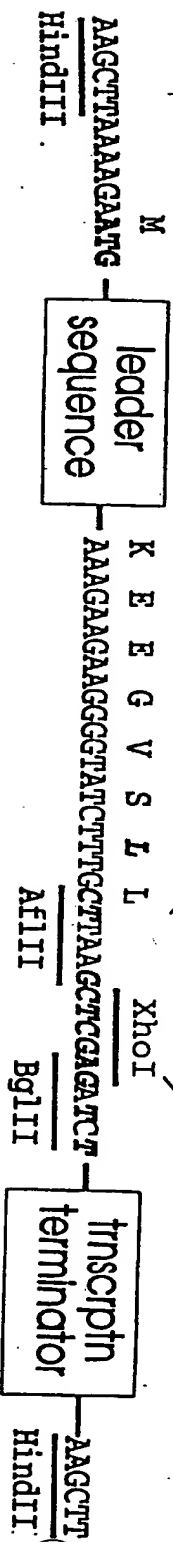
PRINT OF DRAWINGS  
AS ORIGINALLY FILED

FIGURE 1



K R E A E A X<sub>13</sub> \*

CGTGAAGC TTAAGCGTGAGGCAAGACT (NNK)<sub>13</sub> T GATCATCCG  
AflII BclI



ADCI PROMOTER

BstXI  
EagI  
XbaI  
BamHI  
SacI  
SacII  
NotI  
SpeI

HindIII  
Sall  
DraIII  
KpnI  
ClaI  
XhoI  
ApaI

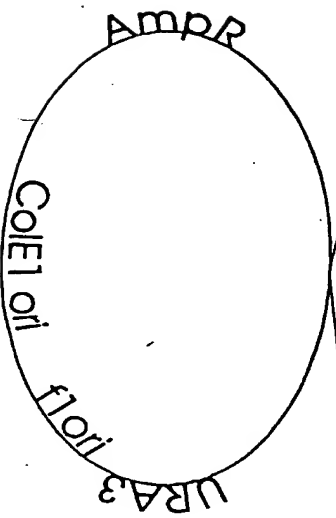


FIGURE 2

S S E K K D N X<sub>11</sub> C V I A \*

GGTAC TCGAGTGAAGAAGAGACAC (NNK)<sub>11</sub> TGTGTATTGCTTAA GTACG

XhoI AflII

SacI

XhoI AflII

AGCTTTTCGAATAGAATG  
HindIII M

Mfcl  
leader

GGCGCTCCAAAAGAAGACCTCGAGCTCGCTTAAG  
A A P K E K T S S

transcript  
terminator

AAGCTT  
HindIII

ADC1 PROMOTER

SacI BstXI  
SacII EagI  
NotI XbaI  
SpeI BamHI  
HindIII  
SalI  
DraIII  
KpnI  
ClaI  
XhoI  
ApaI

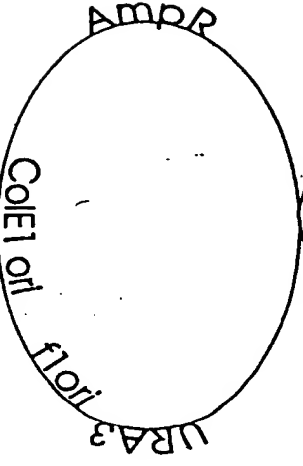
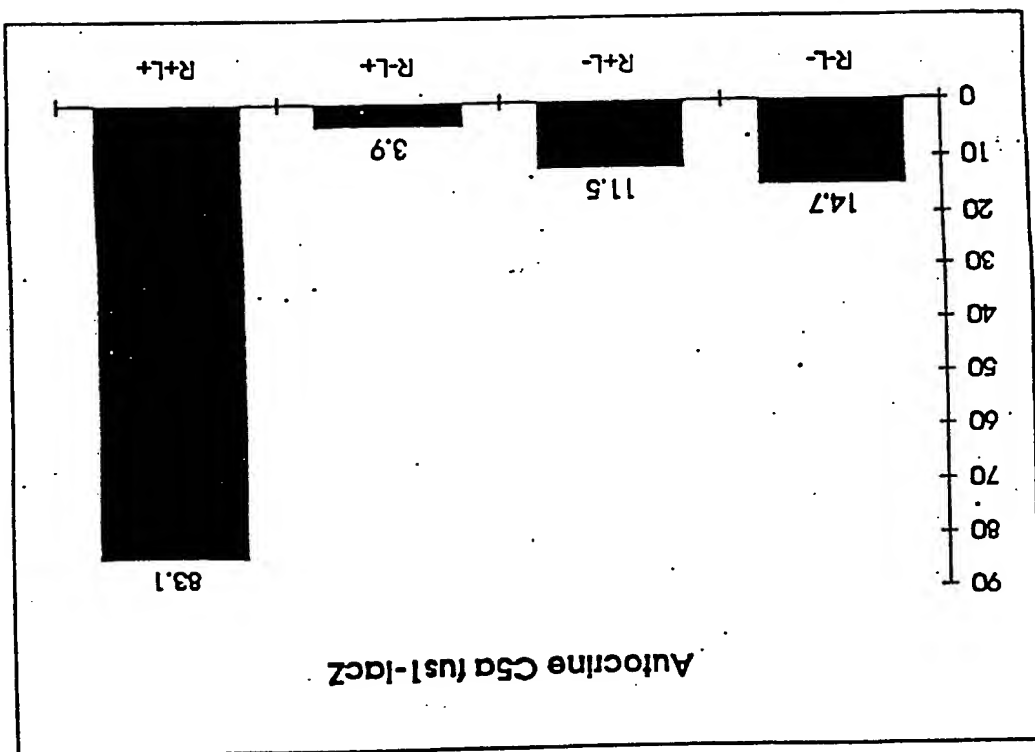


FIGURE 3

08/587895

FIGURE 4



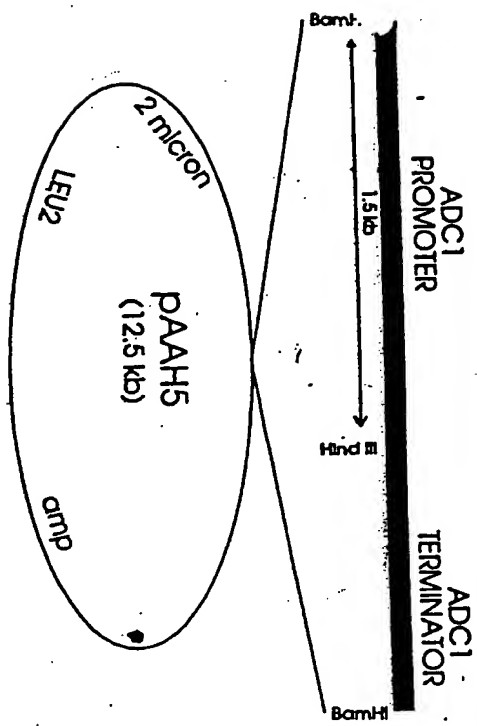
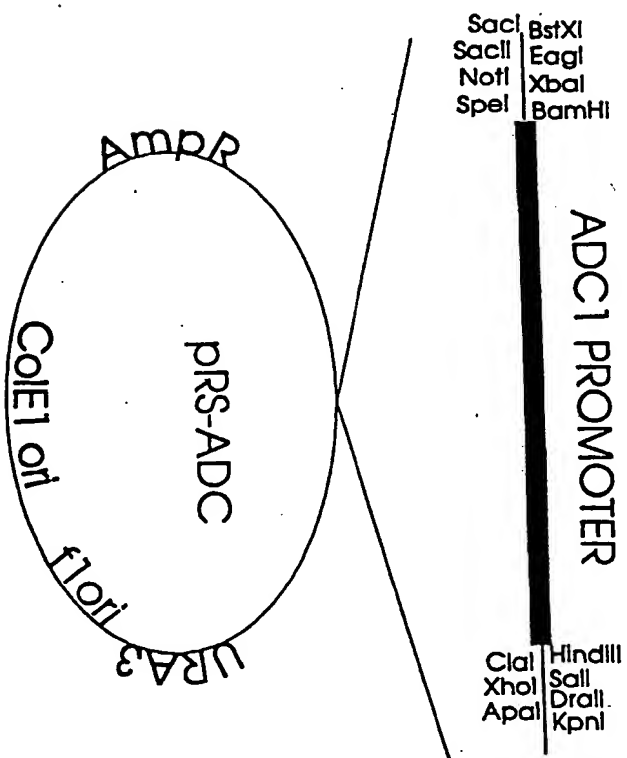


FIGURE 1



K R E A E A X<sub>13</sub> \*

CGTGAAGC TTAAGCGTGAGCGCAGAAGCT (NMK)<sub>13</sub> T GATCATCCG  
AflII BclI

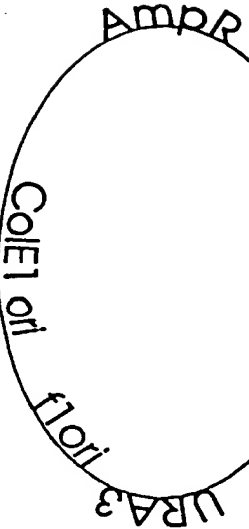
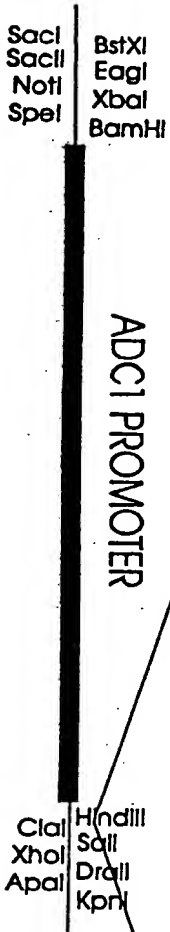
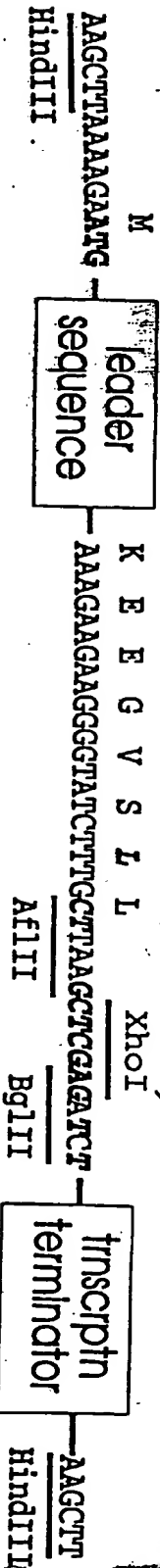
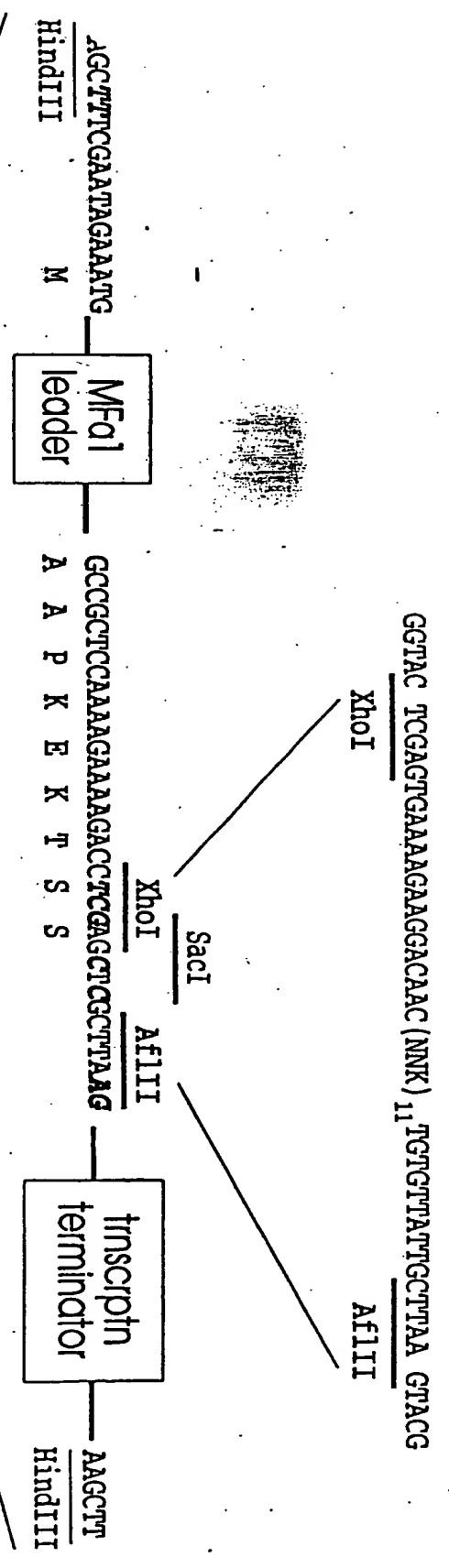
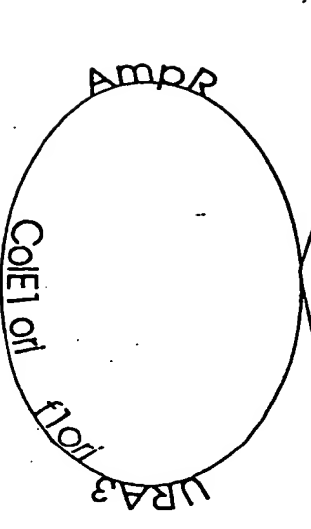


FIGURE 2



S S E K K D N X<sub>11</sub> C V I A \*

GGTAC TCGAGTGAAGAAGACAC (NNK)<sub>11</sub> TGTGTAATGCTTAA GTACG

XhoI AflII

FIGURE 3



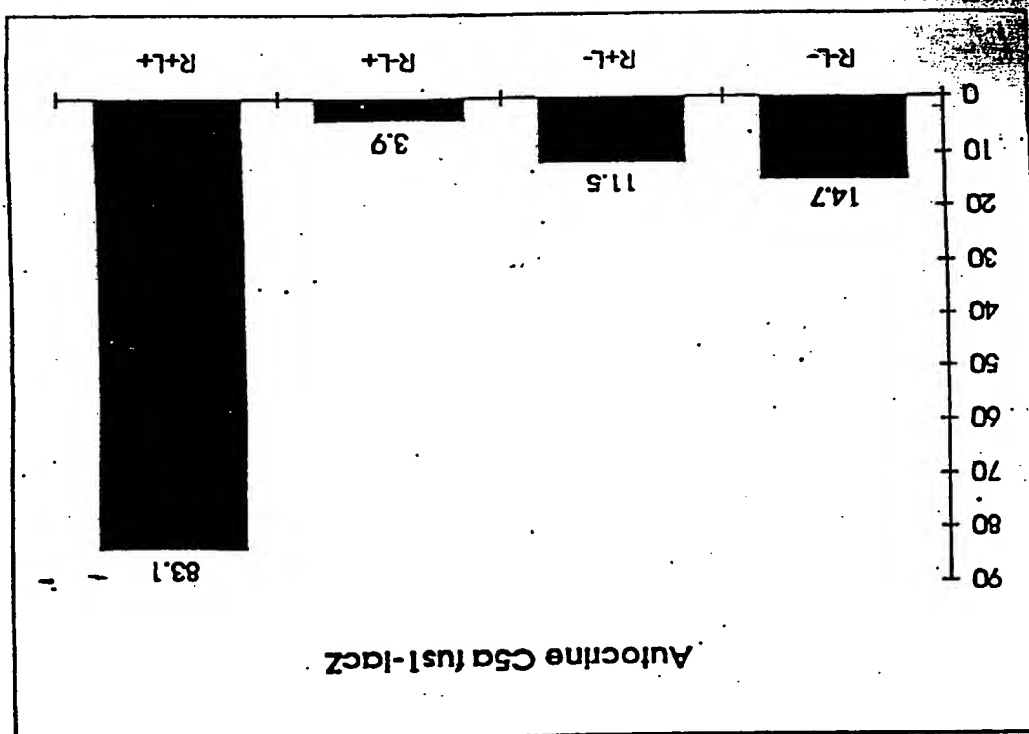
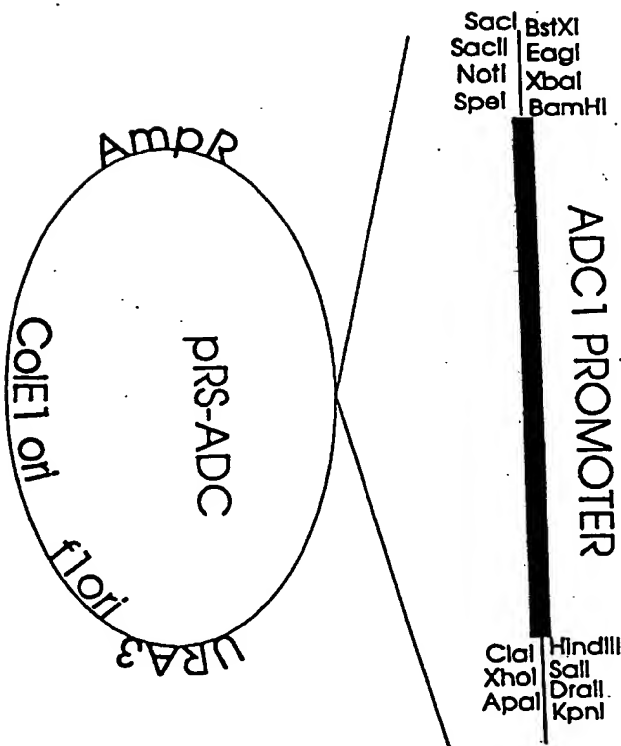
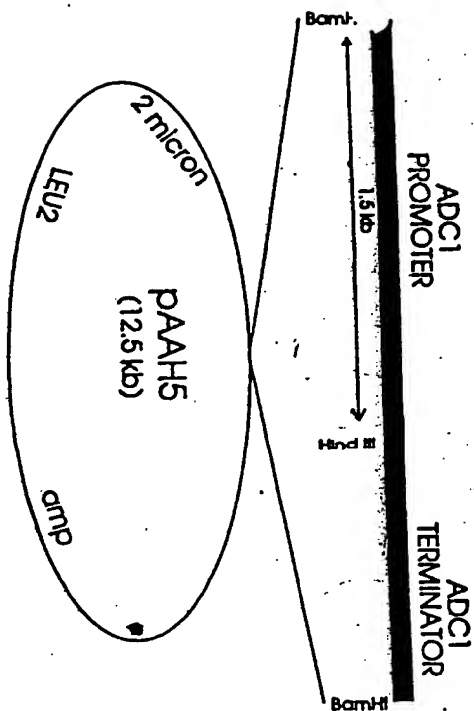


FIGURE 4

08/587895

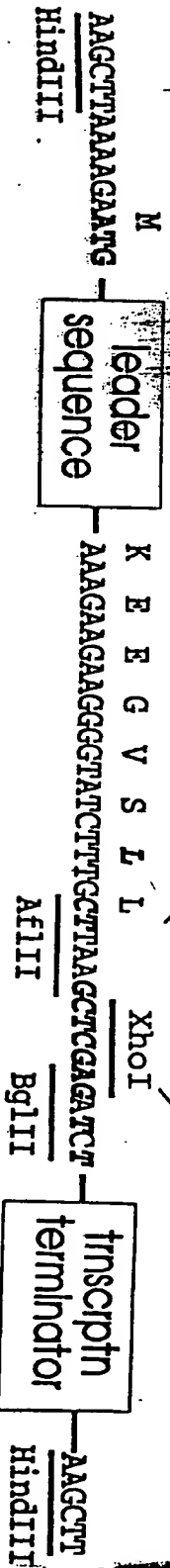
PRINT OF DRAWINGS  
AS ORIGINALLY FILED

FIGURE 1



K R E A E A X<sub>3</sub> \*

CGTGAAGC TTAAGCGTGAGGACAGAGCT (NNK)<sub>13</sub> T GATCATCCG  
AflII BclI



ADC1 PROMOTER

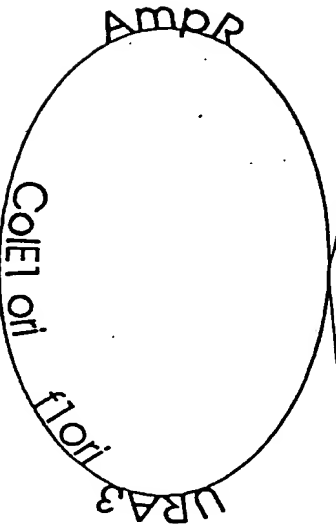


FIGURE 2

S S E K K D N X<sub>11</sub> C V I A \*

GGTAC TCGAGTGAAGAAGACAC(NNK)<sub>11</sub> TGTGTATTGCTTAA GTACG

XhoI

AflII

SacI

XhoI

AflII

AGCTTTCGATAGAATG  
HindIII

M

Mfcl  
leader

GCCGCTCCAAAGAAAGACCCTGAGCTCGCTTAAG  
A A P K E K T S S

transcrip  
terminator

AAGCTT  
HindIII

ADCl PROMOTER

BstXI  
EagI  
XbaI  
BamHI  
SacI  
SacII  
NotI  
SpeI

HindIII  
SalI  
DraII  
KpnI  
ClaI  
XhoI  
ApaI

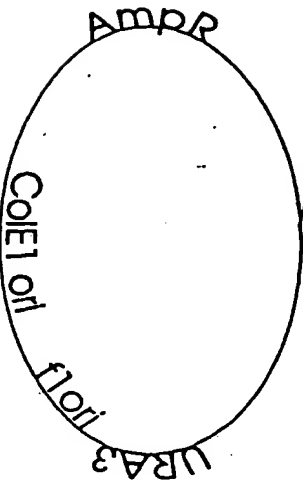


FIGURE 3

S S E K K D N X<sub>11</sub> C V I A \*

GGTAC TCGAGTGAAGAAGAGACAC(NNK)<sub>11</sub>TGTGTATTGCTTAA GTACG

XhoI

AflII

SacI

XhoI

AflII

AGCTTTCGATAGAAATG

M

MFol  
leader

GCCGCTCCAAGAAGAAGACCCTGAGCTCGCTTAAG

A A P K E K T S S

transcript  
terminator

AAGC

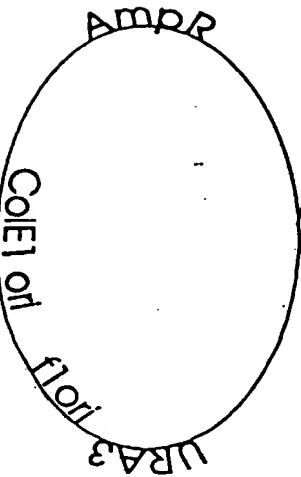
HindIII

ADC1 PROMOTER

BstXI  
EagI  
XbaI  
BamHI  
SacI  
SacII  
NotI  
SpeI

HindIII  
SalI  
DraIII  
KpnI  
ClaI  
XhoI  
ApaI

FIGURE 3



08/587895

FIGURE 4

